

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:59:05 ; Search time 27 seconds

(without alignments)
1496.535 Million cell updates/sec

Title: US-09-874-162A-8

Perfect score: 776

Sequence: 1 MTGIAAGFFSNTCRFGCG.....KALETDSVSGVSKQKKQL 776

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 647 | 83.4 | 739 | 1 JUAZ_HUMAN | Q15022 homo sapien |
| 2 | 272 | 35.1 | 741 | 1 JUAZ_MOUSE | Q80u70 mus musculus |
| 3 | 128 | 16.5 | 243 | 1 JAZ1_HUMAN | Q86vz6 homo sapien |
| 4 | 128 | 16.5 | 243 | 1 JAZ1_MOUSE | Q80zq5 mus musculus |
| 5 | 11 | 1.4 | 900 | 1 SU12_DROME | Q9njg9 drosophila |
| 6 | 8 | 1.0 | 247 | 1 Y124_RPVAC | P41679 autographa |
| 7 | 8 | 1.0 | 662 | 1 GGT4_MOUSE | Q99jp7 mus musculus |
| 8 | 8 | 1.0 | 662 | 1 GGT4_RAT | Q99mz4 rattus norv |
| 9 | 8 | 1.0 | 666 | 1 GGT4_HUMAN | Q9u114 homo sapien |
| 10 | 8 | 1.0 | 669 | 1 COTE_HUMAN | P81408 homo sapien |
| 11 | 8 | 1.0 | 1175 | 1 P1B4_HUMAN | Q15147 homo sapien |
| 12 | 7 | 0.9 | 46 | 1 THN2_VISAL | P32880 viscum albu |
| 13 | 7 | 0.9 | 46 | 1 THN_PHOTO | P01539 phoradendro |
| 14 | 7 | 0.9 | 105 | 1 SUG2_CITFR | Q69279 citrobacter |
| 15 | 7 | 0.9 | 139 | 1 ACPS_BRAJA | Q69159 bradyrhizob |
| 16 | 7 | 0.9 | 139 | 1 ASNC_ECOLI | P03809 escherichia |
| 17 | 7 | 0.9 | 153 | 1 E417_ADE02 | P03238 human adeno |
| 18 | 7 | 0.9 | 154 | 1 NHP2_SCHPO | Q9P7h0 schizosach |
| 19 | 7 | 0.9 | 159 | 1 MUTX_STRMO | P95781 streptococc |
| 20 | 7 | 0.9 | 205 | 1 RUVA_BACAA | Q81l98 bacillus an |
| 21 | 7 | 0.9 | 205 | 1 RUVA_BACCA | Q817w3 bacillus ce |
| 22 | 7 | 0.9 | 234 | 1 POLN_RHDV3 | P27411 rabbit hemo |
| 23 | 7 | 0.9 | 243 | 1 NUKS_RAT | Q9epj0 rattus norv |
| 24 | 7 | 0.9 | 269 | 1 H1S9_LACLA | Q02150 lactococcus |
| 25 | 7 | 0.9 | 269 | 1 TH1D_STRCO | Q9zbr6 streptomyce |
| 26 | 7 | 0.9 | 274 | 1 ABCD_HUMAN | Q9nse7 homo sapien |
| 27 | 7 | 0.9 | 293 | 1 ERA_MYCPU | Q98q11 mycoplasma |
| 28 | 7 | 0.9 | 294 | 1 E434_ADE02 | P03239 human adeno |
| 29 | 7 | 0.9 | 319 | 1 YC39_ODOSI | P49534 odontella s |
| 30 | 7 | 0.9 | 324 | 1 ANXA_HUMAN | Q9ui72 homo sapien |
| 31 | 7 | 0.9 | 329 | 1 YG54_ARCFU | Q28619 archaeoglob |
| 32 | 7 | 0.9 | 368 | 1 SH31_HUMAN | Q99961 homo sapien |
| 33 | 7 | 0.9 | 368 | 1 SH31_MOUSE | Q62419 mus musculus |

ALIGNMENTS

RESULT 1

```
JUAZ_HUMAN
ID JUAZ_HUMAN STANDARD; PRT; 739 AA.
AC Q15022; Q96BD9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Polycarb protein SUZ12 (Suppressor of zeste 12 protein homolog)
DE (Joined to JAZF1 protein) (Chromatin precipitated E2F target 9
DE protein) (ChET 9 protein).
GN SUZ12 OR JJA21 OR CHET9 OR KIAA0160.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RA MEDLINE=96127530; PubMed=8590280;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:167-174(1995).
RX [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RX [3]
RP SEQUENCE OF 102-113; 343-360; 474-483 AND 509-516, IDENTIFICATION IN A
RP PR2 COMPLEX WITH E2F; E2H2; RBP4 AND RBBP7, AND METHYLTRANSFERASE
RP ACTIVITY OF THE COMPLEX.
RX MEDLINE=22300550; PubMed=12351676;
RA Cao R., Wang L., Wang H., Xia L., Erdjument-Bromage H., Tempst P.,
RA Jones R.S., Zhang X.;
RT "Role of histone H3 lysine 27 methylation in Polycomb-group
```

P49177 arabidopsis
Q56806 xanthomonas
P72242 pseudomonas
Q926m3 chlamydia p
Q59671 pseudomonas
Q60140 pseudomonas
Q8ue16 agrobacteri
P75022 agrobacteri
Q8yh04 brucella me
Q925y6 rhizobium m
P42474 cytophaga l
Q98qp8 pseudomonas

silencing.";
 RL Science 298:1039-1043 (2002).
 [4]
 RN DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH JAZF1.
 RX MEDLINE=21265470; PubMed=11371647;
 RA Krontz J.I., Soreng A.L., Nucci M., Kuo F.C., Pauwels P.,
 Van Den Berghe H., Clin P.D., Fletcher J.A., Sklar J.;
 RT "Frequent fusion of the JAZF1 and JUAZ1 genes in endometrial stromal
 tumors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353 (2001).
 [5]
 RN INDUCTION.
 RP MEDLINE=21448978; PubMed=11564866;
 RA Weimann A.S., Bartley S.M., Zhang T., Zhang M.Q., Farnham P.J.;
 RT "Use of chromatin immunoprecipitation to clone novel E2F target
 promoters.";
 RL Mol. Cell. Biol. 21:6820-6832 (2001).
 [6]
 RN IDENTIFICATION IN A PRC2 COMPLEX WITH EED; EZH2; RBBP4 AND RBBP7, AND
 RP METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
 RX MEDLINE=22323099; PubMed=12435631;
 RA Kuzmichev A., Nishio K., Erdjument-Bromage H., Tempst P.,
 Reinberg D.;
 RT "Histone methyltransferase activity associated with a human
 multiprotein complex containing the Enhancer of Zeste protein.";
 RL Genes Dev. 16:2893-2905 (2002)
 CC -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by
 forming multiprotein complexes, which are required to maintain the
 transcriptionally repressive state of homeotic genes throughout
 development. PcG proteins are not required to initiate repression,
 but to maintain it during later stages of development. They
 probably act via the methylation of histones, rendering chromatin
 heritably changed in its expressibility. Component of the PRC2
 complex, which methylates Lys-9 and Lys-27 residues of histone H3.
 CC -!- SUBUNIT: Component of the PRC2 complex, composed of EED, EZH2,
 SUZ12/JAZ1, RBBP7. This complex is distinct from the
 PRC1 complex, which contains many other PcG proteins like BMI1,
 CBX4, CBX6, PHC3, SCMH1, RING1 and RNF2.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- INDUCTION: Induced by E2F1 transcription factor.
 CC -!- DISEASE: Endometrial stromal tumors can be caused by a chromosomal
 rearrangement involving JAZF1 and SUZ12. A t(7;17)(p15;q21)
 chromosomal translocation involving JAZF1 and SUZ12 generates the
 JAZF1-SUZ12 oncogene consisting of the N-terminus part of JAZF1
 and the C-terminus part of SUZ12. The translocation is frequently
 found in all cases of endometrial stromal tumors, except in
 endometrial stromal sarcomas where it is rarer.
 CC -!- SIMILARITY: Belongs to the VEFs (VRN2-EMF2-FIS2-SU(2)12) family.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 or send an email to license@sib-sib.ch).

 DR EMBL; D63881; BAA09931.1; ALT INIT.
 DR EMBL; BC015704; AAH15704.1; -.
 DR MIM; 606245; -.
 DR InterPro; IPR007087; Znf_C2H2.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE NEG.
 KW Transcription regulation; Repressor; Chromatin regulator;
 KW Nuclear protein; Metal-binding; Zinc; Zinc-finger;
 KW Chromosomal translocation; Proto-oncogene.
 FT ZN FING 448 471 C2H2-TYPE.
 FT DOMAIN 563 639 VEFs-BOX.
 FT DOMAIN 7 50 GLY-RICH.
 FT DOMAIN 51 59 POLY-SER.
 FT DOMAIN 60 67 POLY-ALA.
 FT SITE 93 94 BREAKPOINT FOR TRANSLOCATION TO FORM

FT JAZF1-SUZ12 ONCOGENE.
 SQ SEQUENCE 739 AA; 83053 MW; A8830EBD031DF656 CRC64;
 Query Match 83.4%; Score 647; DB 1; Length 739;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 PTQIVRFLTRNLIAPIFLHRTLTYNSHRNSTNTKRTFKVDDMLSKVKMKQGEQSHS 189
 DB 93 PTQIVRFLTRNLIAPIFLHRTLTYNSHRNSTNTKRTFKVDDMLSKVKMKQGEQSHS 152
 QY 190 LSAHQLQTLTGFFHNDKPSNSNEQNSQVTLVLLVVKVCHKKRDXSCPIRQVPTGKKQ 249
 DB 153 LSAHQLQTLTGFFHNDKPSNSNEQNSQVTLVLLVVKVCHKKRDXSCPIRQVPTGKKQ 212
 QY 250 VPLIPDLNQTGPNFSLAVSSNEPEPNSHMKVSYSLLFVTRPGRREFNGMINGETNE 309
 DB 213 VPLIPDLNQTGPNFSLAVSSNEPEPNSHMKVSYSLLFVTRPGRREFNGMINGETNE 272
 QY 310 NIDVNEELPARRKRNREDGEKTFVAQMTVFDKRNRLQLLDGEYEVAMQEMEECPISKRA 369
 DB 273 NIDVNEELPARRKRNREDGEKTFVAQMTVFDKRNRLQLLDGEYEVAMQEMEECPISKRA 332
 QY 370 TWETILDGKRLPPPTFFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSSLHQENKPG 429
 DB 333 TWETILDGKRLPPPTFFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSSLHQENKPG 392
 QY 430 SVKPTQTIKVESLTDTLOTRKEKTPNENROKLIKFQFLYNNTRQOTEARDDLHCPW 489
 DB 393 SVKPTQTIKVESLTDTLOTRKEKTPNENROKLIKFQFLYNNTRQOTEARDDLHCPW 452
 QY 490 CTLNCRKLYSLKHLKCHSRFIPNVYHPKGRDVSINECYDGSYAGNPQDIHQPGF 549
 DB 453 CTLNCRKLYSLKHLKCHSRFIPNVYHPKGRDVSINECYDGSYAGNPQDIHQPGF 512
 QY 550 AFSRNGPVKRTPIHLVCRPKRTKASMEFLESDGEVEQRTYSSGHNRLYFHSDFCL 609
 DB 513 AFSRNGPVKRTPIHLVCRPKRTKASMEFLESDGEVEQRTYSSGHNRLYFHSDFCL 572
 QY 610 PLRQMEVDSDEKDPWLREKTIQTEESDVNNEGEKVMKLNHLVHMKHGFADNQ 669
 DB 573 PLRQMEVDSDEKDPWLREKTIQTEESDVNNEGEKVMKLNHLVHMKHGFADNQ 632
 QY 670 NEACMLFVENYQKIKKNLCNFMHLVSMHDFNLISIMSIDKAVTKLREMQQKLEGE 729
 DB 633 NEACMLFVENYQKIKKNLCNFMHLVSMHDFNLISIMSIDKAVTKLREMQQKLEGE 692
 QY 730 SASPANEETEEQNTANGFSEINSKEKALETDSVSGVSKSKQKQL 776
 DB 693 SASPANEETEEQNTANGFSEINSKEKALETDSVSGVSKSKQKQL 739
 RESULT 2
 ID JUAZ MOUSE STANDARD; PRT; 741 AA.
 AC Q80U70; Q80Y10; Q99L07;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Polycomb protein Suz12 (Suppressor of zeste 12 protein homolog).
 GN SUZ12 OR KIAA0160 OR D11ERTD530E.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
 Nakajima D., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:

RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
 RT cDNAs identified by screening of terminal sequences of cDNA clones
 RT randomly sampled from size-fractionated libraries."
 RL DNA Res. 10:35-48(2003).
 RN [2]
 RP SEQUENCE OF 365-741 FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Polycomb group (PCG) protein. PCG proteins act by
 CC forming multiprotein complexes, which are required to maintain the
 CC transcriptionally repressive state of homeotic genes throughout
 CC development. PCG proteins are not required to initiate repression,
 CC but to maintain it during later stages of development. They
 CC probably act via the methylation of histones, rendering chromatin
 CC heritably changed in its expressibility. Component of the PRC2
 CC complex, which methylates Lys-9 and Lys-27 residues of histone H3
 CC (By similarity).
 CC -!- SUBUNIT: Component of the PRC2 complex, composed of EED, EZH2,
 CC SUZ12/JAZ1, RBBP4 and RBBP7. This complex is distinct from the
 CC PRC1 complex, which contains many other PCG proteins like EML1,
 CC CBX4, CBX6, PRC2, PRC3, SCMH1, RING1 and RNF2 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Belongs to the VEFs (VRN2-EMF2-FIS2-SU(2)12) family.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to an
 CC erroneous RNA joining and initiations.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AK122213; BAC65495.1; ALT_SEQ.
 DR EMBL; BC003922; AAH03922.1; -.
 DR EMBL; BC051099; AAH51099.1; -.
 DR MGD; MGI:1261758; D11Etd530e.
 DR InterPro; IPR007087; Znf_C2H2.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; FALSE NEG.
 KW Transcription regulation; Repressor; Chromatin regulator;
 KW Nuclear protein; Metal-binding; Zinc; Zinc-finger.
 FT ZN FING 450 473 C2H2-TYPE.
 FT DOMAIN 565 641 VEFs-BOX.
 FT DOMAIN 7 50 GLY-RICH.
 FT DOMAIN 29 63 AUA-RICH.
 SQ SEQUENCE 741 AA; 83025 MW; 4ACB6E52D24FAEF2 CRC64;
 Query Match 35.1%; Score 272; DB 1; Length 741;
 Best Local Similarity 100.0%; Pred. No. 7.5e-276;
 Matches 272; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 460 RQKLRIYQFLYNNNTQQTEARDLHCPWCTLCNCRKLYSLKHLKLCRSRFFNYVYHP 519
 DB 425 RQKLRIYQFLYNNNTQQTEARDLHCPWCTLCNCRKLYSLKHLKLCRSRFFNYVYHP 484
 QY 520 KGARIDVINECYDGSVAGNPQDIHQPGFAPGRNGPVKRTPTTHILVCPKPKTKASMSE 579
 DB 485 KGARIDVINECYDGSVAGNPQDIHQPGFAPGRNGPVKRTPTTHILVCPKPKTKASMSE 544
 QY 580 FLESEDEVEQQRYSYSSGNRLYFHSDTCLPLRPOEVEVSEDEKDPWELREKTIITQIEE 639
 DB 545 FLESEDEVEQQRYSYSSGNRLYFHSDTCLPLRPOEVEVSEDEKDPWELREKTIITQIEE 604
 QY 640 FSDVNEGEKEVVKLWNLHVNMKGFIADNQNHACMLFVENYQKIKKNCNRMFLHVS 699
 DB 605 FSDVNEGEKEVVKLWNLHVNMKGFIADNQNHACMLFVENYQKIKKNCNRMFLHVS 664
 QY 700 MHDNPLNISIMSDKAVTKLREMOQKLEKGESA 731
 DB 665 MHDNPLNISIMSDKAVTKLREMOQKLEKGESA 696
 RESULT 3
 JAZ1 HUMAN
 ID JAZ1 HUMAN STANDARD; PRT; 243 AA.
 AC Q86VZ6; G8N3L7;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Juxtaposed with another zinc finger protein 1.
 GN JAZF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain, and Testis;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Amnygdala;
 RX Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RA [3]
 RP DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH SUZ12.
 RX MEDLINE=21265470; PubMed=11371647;
 RA Koonz J.I., Soreng A.L., Nucci M., Kuo F.C., Pauwels P.,
 RA van Den Berghe H., Cin P.D., Fletcher J.A., Sklar J.;
 RT "Frequent fusion of the JAZF1 and JAZ1 genes in endometrial stromal
 RL tumors."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353(2001).
 CC -!- FUNCTION: Potential transcription factor.

CC -!- SUBCELLULAR LOCATION: Nuclear protein (Potential).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q86VZ6-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q86VZ6-2; Sequence=VSP_007755;
 CC Note=Sequence incomplete. No experimental confirmation available;
 CC -!- DISBAS: Endometrial stromal tumors can be caused by a chromosomal rearrangement involving JAZF1 and SUZ12. A t(7;17)(p15;q21) chromosomal translocation involving JAZF1 and SUZ12 generates the JAZF1-SUZ12 oncogene consisting of the N-terminus part of JAZF1 and the C-terminus part of SUZ12. The translocation is frequently found in all cases of endometrial stromal tumors, except in endometrial stromal sarcomas, where it is rarer.
 CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
 CC -!- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; BC042441; AAH42441.1; -;
 CC EMBL; AL834234; CAD38912.1; -;
 CC MIM; 606246; -;
 CC InterPro; IPR007087; Znf C2H2.
 CC Pfam; PF00096; Zf-C2H2; 3.
 CC SMART; SM00355; Znf C2H2; 3.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_2; FALSE NEG.
 CC Transcription regulation; Nuclear protein; Metal-binding; Zinc;
 CC Zinc-finger; Chromosomal translocation; Proto-oncogene;
 CC Alternative splicing.
 CC FT ZN_FING 12 37 C2H2-TYPE 1.
 CC FT ZN_FING 173 198 C2H2-TYPE 2.
 CC FT ZN_FING 208 230 C2H2-TYPE 3 (DEGENERATE).
 CC FT SITE 129 130 BREAKPOINT FOR TRANSLLOCATION TO FORM JAZF1-SUZ12 ONCOGENE.
 CC FT VARSPLIC 1 38 MTGIAAASFNTCTFCGCGHFFPTLADLIEHIDNHDTPRVLEKQELQPTVVALSY ->
 CC FT YFVLGSLIYSDAKSKNNRCVFIT (in isoform 2).
 CC FT /FTID=VSP_007755.
 CC FT CONFLICT 133 133 D -> G (IN REF. 1).
 CC SEQUENCE 243 AA; 27079 MW; 878A6E3D82C0588 CRC64;
 CC Query Match 16.5%; Score 128; DB 1; Length 243;
 CC Best Local Similarity 100.0%; Pred. No. 1.2e-125;
 CC Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTGIAAASFNTCTFCGCGHFFPTLADLIEHIDNHDTPRVLEKQELQPTVVALSY 60
 DB 1 MTGIAAASFNTCTFCGCGHFFPTLADLIEHIDNHDTPRVLEKQELQPTVVALSY 60
 QY 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSGNVSSTPPRHSSGSLTPPTPTPSS 120
 DB 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSGNVSSTPPRHSSGSLTPPTPTPSS 120
 QY 121 SFRSSTPT 128
 DB 121 SFRSSTPT 128
 RESULT 4
 JAZ1_MOUSE
 ID JAZ1_MOUSE STANDARD; PRT; 243 AA.
 AC Q80ZQ5; Q8BLQ8; Q8BNW4;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Juxtaposed with another zinc finger protein 1.

GN JAZF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/6J; TISSUE=Brain cortex, and Pancreas;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaido I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Mikhi H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Pavlovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Sempole C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RT Nature 420:563-573(2002).
 RL [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Testicle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grumman J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Potential transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear protein (Potential).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q80ZQ5-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q80ZQ5-2; Sequence=VSP_007756;
 CC Note=No experimental confirmation available;
 CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
 CC -----
 CC

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)

CC -----
 CC EMBL: AK043767; BAC31649.1; -
 CC EMBL: AK050474; BAC34275.1; -
 CC EMBL: BC048577; AAH48577.1; -
 CC InterPro: IPR007087; Znf_C2H2.
 CC Pfam: PF00096; zf-C2H2; 3.
 CC SMART: SW00355; Znf_C2H2; 3.
 CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
 CC PROSITE: PS0157; ZINC_FINGER_C2H2_2; FALSE NEG.
 CC Transcription regulation; Nuclear protein; Metal-binding; Zinc;
 CC Zinc-finger; Alternative splicing
 CC ZN_FING 12 37 C2H2-TYPE 1.
 CC ZN_FING 173 198 C2H2-TYPE 2.
 CC ZN_FING 208 230 C2H2-TYPE 3 (DEGENERATE).
 CC FT VARSPLIC 64 64 Missing (in isoform 2).
 CC FT FTID=VSP_007756.
 CC FT CONFLICT 211 211 R -> L (IN REF. 1; BAC34275).
 CC SQ SEQUENCE 243 AA; 27097 MW; 878A6EE3D82C0348 CRC64;

Query Match 16.5%; Score 128; DB 1; Length 243;
 Best Local Similarity 100.0%; Pred. No. 1.2e-125;
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFSTCRFGGGLHPFTLADLIEHNHDTDRVLEKELQOPTVALSY 60
 DB 1 MTGIAAASFSTCRFGGGLHPFTLADLIEHNHDTDRVLEKELQOPTVALSY 60
 QY 61 INRFMTDAARRQESLKKIQKLSLTSSSVSGNVSTPPRHSGSLTPPTPTPSS 120
 DB 61 INRFMTDAARRQESLKKIQKLSLTSSSVSGNVSTPPRHSGSLTPPTPTPSS 120
 QY 121 SFRSSTPT 128
 DB 121 SFRSSTPT 128

RESULT 5

ID_SUI2_DROME STANDARD; PRT; 900 AA.
 AC Q9NKG9; Q879D8; Q9VW55;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Polycomb protein Su(z)12 (Suppressor 12 of zeste protein).
 GN SU(z)12 OR CG8013.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), CHARACTERIZATION, AND MUTANT
 RP SU(z)12-2.
 RX MEDLINE=21430867; PubMed=11546753;
 RA Birve A., Sengupta A.K., Beuchle D., Larsson J., Kennison J.A.,
 RA Rasmuson-Lestander A., Mueller J.;
 RT "Su(z)12, a novel Drosophila Polycomb group gene that is conserved in
 RT vertebrates and plants";
 RL Development 128:3371-3379 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Morkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195 (2000).
 RN [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX STRAIN=Berkeley; TISSUE=Embryo;
 MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A Drosophila full-length cDNA resource";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
 RN [5]
 RP IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAP1 AND ESC,
 RP AND METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
 RX MEDLINE=22296673; PubMed=12408863;
 RA Czermin B., Meli R., McCabe D., Seitz V., Imhof A., Pirrotta V.;
 RT "Drosophila enhancer of Zeste/ESC complexes have a histone H3
 RT methyltransferase activity that marks chromosomal Polycomb sites";
 RL Cell 111:185-196 (2002).
 RN [6]
 RP IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAP1 AND ESC, AND
 RP METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
 RX MEDLINE=22296674; PubMed=12408864;
 RA Mueller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A.,
 RA Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A.;
 RT "histone methyltransferase activity of a Drosophila Polycomb group
 RT repressor complex.";

NCBI_TaxID=10090;
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nagai K., Tomaru Y., Hasegawa Y., Nogami A., Schombach C., Gofjohori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Flecher C.F., Forrest A., Frazer K.S.,
RA Gascierland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hircokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedziński R.M., King B.L.,
RA Kottagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed J.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wiland L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Koshikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast cancer;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE OF 552-562 FROM N.A.
RC STRAIN=ISS, and ILS;
RX MEDLINE=21363810; PubMed=11471062;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-throughput sequence identification of gene coding variants
within alcohol-related QTLs.";
RL Mamm. Genome 12:657-663 (2001).
CC -!- FUNCTION: Cleaves glutathione conjugates (By similarity).
CC peptide + 5-L-glutamyl-amino acid.
CC -!- PATHWAY: Glutathione metabolism.
CC -!- SUBUNIT: Heterodimer composed of the light and heavy chains. The
active site is located in the light chain. Interacts with CTL20

(By similarity).
-!- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
-!- SIMILARITY: Belongs to the gamma-glutamyltransferase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AK032051; BAC27672.1; -
DR EMBL; BC005772; AAH05772.1; -
DR EMBL; AF332053; AAK56082.1; -
DR EMBL; AF332054; AAK56083.1; -
DR MGD; MGI:1913385; Ggt13.
DR InterPro; IPR000101; Peptidase_T3.
DR Pfam; PF01019; G_glu_transf_1.
DR PRINTS; PR01210; GGTRANSFERASE.
DR PROSITE; PS00462; G_GLU_TRANSFERASE; FALSE NEG.
KW Glutathione biosynthesis; Transferase; Acyltransferase; Signal-anchor;
KW Transmembrane; Zymogen; Glycoprotein.
FT CHAIN 1 472
FT CHAIN GAMMA-GLUTAMYLTRANSFERASE 4 HEAVY
FT CHAIN (BY SIMILARITY).
FT CHAIN GAMMA-GLUTAMYLTRANSFERASE 4 LIGHT
FT DOMAIN 1 106
FT TRANSMEM 107 127
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 128 662
FT CARBOHYD 198 198 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 283 283 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 523 523 N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC.) (POTENTIAL).
FT CONFLICT 148 148 H -> S (IN REF. 1).
SQ SEQUENCE 662 AA; 70301 MW; BA7F32512CC4F381 CRC64;
Query Match 1.0%; Score 8; DB 1; Length 662;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 172 DMLSKVE 179
Db 436 DMLSKVE 443

RESULT 8
GGT4 RAT
ID GGT4 RAT STANDARD; PRT; 662 AA.
AC Q99MZ4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Gamma-glutamyltransferase 4 precursor (EC 2.3.2.2) (Gamma-
DE glutamyltransferase 4) (Gamma-glutamyltransferase-like 3).
GN GGT13.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 3-662 FROM N.A.
RA Yamaguchi T., Araki K., Nawa H.;
RT "Rattus norvegicus gamma-glutamyltransferase homolog mRNA complete
code.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cleaves glutathione conjugates (By similarity).

CC -!- CATALYTIC ACTIVITY: (5-L-glutamyl)-peptide + an amino acid =
 CC peptide + 5-L-glutamyl-amino acid.
 CC -!- PATHWAY: Glutathione metabolism.
 CC -!- SUBUNIT: Heterodimer composed of the light and heavy chains. The
 CC active site is located in the light chain. Interacts with CT120
 CC (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
 CC -!- SIMILARITY: Belongs to the gamma-glutamyltransferase family.
 CC -!- CAUTION: The N-terminus was extended based on the genomic
 CC sequence, in analogy to ortholog sequences.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: AF244973; AAK27971.1; ALT INIT.
 CC InterPro: IPR000101; Peptidase_F3.
 CC Pfam: PF01019; G-glu_transferpep; 1.
 CC PRINTS: PR01210; GGTRANSFERASE.
 CC PROSITE: PS00462; G_GLU_TRANSFERPEPTIDASE; FALSE NEG.
 CC KX Glutathione biosynthesis; Transferase; Acyltransferase; Signal-anchor;
 CC Transmembrane; Zymogen; Glycoprotein.
 CC FT CHAIN 1 472
 CC GAMMA-GLUTAMYLTRANSFERASE 4 HEAVY
 CC
 CC FT CHAIN 473 662
 CC GAMMA-GLUTAMYLTRANSFERASE 4 LIGHT
 CC CHAIN (BY SIMILARITY).
 CC FT DOMAIN 1 106
 CC CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 107 127
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC FT DOMAIN 128 562
 CC EXTRACELLULAR (POTENTIAL).
 CC FT CARBOHYD 198 198
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 267 267
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 283 283
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 330 330
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 353 353
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 394 394
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 519 519
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 523 523
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 566 566
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SQ SEQUENCE 662 AA; 70382 MW; F98B7643700BB03A CRC64;
 CC
 CC Query Match 1.0%; Score 8; DB 1; Length 662;
 CC Best Local Similarity 100.0%; Pred. No. 10;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC OY 172 DDMLSKVE 179
 CC |||||
 CC Db 436 DDMLSKVE 443
 CC
 CC RESULT 9
 CC GGT4 HUMAN
 CC ID GGT4 HUMAN STANDARD; PRT; 666 AA.
 CC AC Q9UJ14; Q8N899; Q8NF66; Q9BYP5; Q9BYP6;
 CC DT 10-OCT-2003 (Rel. 42, Created)
 CC DT 10-OCT-2003 (Rel. 42, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Gamma-glutamyltransferase 4 precursor (EC 2.3.2.2) (Gamma-
 CC DE glutamyltransferase 4) (Gamma-glutamyltransferase-like 3)
 CC GN GGT4 OR GGT4S
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
 CC OX NCBI_TaxID=9606;
 CC RN [1]
 CC RP SEQUENCE FROM N.A. (ISOFORM 5), AND INTERACTION WITH CT120.
 CC RX MEDLINE=22230983; PubMed=12270127;
 CC He X., Di Y., Li J., Xie Y., Tang F., Wei L., Zhang Y.,

RA Qin W., Huo K., Li Y., Wan D., Gu J.;
 RA "Molecular cloning and characterization of CT120, a novel
 RA membrane-associated gene involved in amino acid transport and
 RA glutathione metabolism.";
 RA Biochem. Biophys. Res. Commun. 297:528-536(2002).
 RN [2]
 RN SEQUENCE FROM N.A. (ISOFORM 4).
 RN RP
 RN TISSUE=Spleen;
 RN Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
 RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
 RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
 RA "NEDO human cDNA sequencing project.";
 RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A. (ISOFORM 6).
 RN RP
 RN TISSUE=Spleen;
 RN MEDLINE=22579292; PubMed=12693554;
 RA Jikuya H., Takano J., Kikuno R., Hirose M., Nagase T., Nomura N.,
 RA Chara O.;
 RA "Characterization of long cDNA clones from human adult spleen. II. The
 RA complete sequences of 81 cDNA clones.";
 RA DNA Res. 10:49-57(2003).
 RN [4]
 RN SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RN RP
 RN MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.I.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark A.P., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leharshain M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie I.J., McLeay K., Nickerson T.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RA "The DNA sequence and comparative analysis of human chromosome 20.";
 RA Nature 414:865-871(2001).
 CC -!- FUNCTION: Cleaves glutathione conjugates (By similarity).
 CC -!- CATALYTIC ACTIVITY: (5-L-glutamyl)-peptide + an amino acid =
 CC peptide + 5-L-glutamyl-amino acid.
 CC -!- PATHWAY: Glutathione metabolism.
 CC -!- SUBUNIT: Heterodimer composed of the light and heavy chains. The
 CC active site is located in the light chain (By similarity).
 CC Isoform 5 interacts with CT120.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=6;
 CC Name=1; Synonyms=a;
 CC Note=Gene prediction confirmed by EST data;
 CC Name=2; Synonyms=b;
 CC Note=Gene prediction confirmed by EST data;
 CC Name=3; Synonyms=c;

DB 484 GESASPAN 491

RESULT 12

THN2_VISAL STANDARD; PRT; 46 AA.

AC P32880; P01536;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Viscotoxin A2.

GN TH12.3.

OS Viscum album (European mistletoe).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Santalales; Santalaceae; Viscum.

OX NCBI_TaxID=3972;

RN [1]_SEQUENCE.

RP MEDLINE=72211843; PubMed=5035954;

RX Olson T., Samuelsson G.;

RA "The amino acid sequence of viscotoxin A2 from the European mistletoe (Viscum album L., Loranthaceae).";

RL Acta Chem. Scand. 26:585-595(1972).

RN [2]

RP DISULFIDE BONDS.

RX MEDLINE=75013480; PubMed=4607177;

RA Olson T., Samuelsson G.;

RT "The disulphide bonds of viscotoxin A2 from the European mistletoe (Viscum album L. Loranthaceae).";

RL Acta Pharm. Suec. 11:381-386(1974).

CC -!- FUNCTION: Thionins are small plant proteins which are toxic to animal cells. They seem to exert their toxic effect at the level of the cell membrane. The precise function, in plants, of these proteins is not known.

CC -!- SIMILARITY: Belongs to the plant thionin family.

DR PIR; A90005; VTVA2.

DR HSSP; P01544; 2PLH.

DR InterPro; IPR001010; Thionin.

DR Pfam; PF00321; plant thionins; 1.

DR PROSITE; PS00271; THIONIN; 1.

KW Plant defense; Thionin; Plant toxin.

FT DISULFID 3 40 PROBABLE.

FT DISULFID 4 32

FT DISULFID 16 26

SQ SEQUENCE 46 AA; 4834 MW; 61149207EE3FF0E5 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 46;

Best Local Similarity 100.0%; Pred. No. 9.9;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 NTCRFGG 18

DB 14 NTCRFGG 20

RESULT 13

THN_PHOTO STANDARD; PRT; 46 AA.

AC P01539;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phoratoxin.

OS Phoradendron tomentosum (California mistletoe).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Santalales; Santalaceae; Phoradendron.

OX NCBI_TaxID=3969;

RN [1]_SEQUENCE.

RP STRAIN-Subsp. macrophyllum;

RC MEDLINE=75013484; PubMed=4415051;

RX Mellstrand S.T., Samuelsson G.;

RA "Phoratoxin, a toxic protein from the mistletoe Phoradendron tomentosum subsp. macrophyllum (Loranthaceae). The amino acid sequence.";

RL Acta Pharm. Suec. 11:347-360(1974).

RN [2]

RP SEQUENCE.

RA Samuelsson G.;

RT "Mistletoe toxins.";

RL Syst. Zool. 22:566-569(1973).

RN [3]

RP DISULFIDE BONDS.

RX MEDLINE=75013480; PubMed=4606908;

RA Mellstrand S.T., Samuelsson G.;

RT "Phoratoxin, a toxic protein from the mistletoe Phoradendron tomentosum subsp. macrophyllum (Loranthaceae). The disulphide bonds.";

RL Acta Pharm. Suec. 11:367-374(1974).

CC -!- FUNCTION: Thionins are small plant proteins which are toxic to animal cells. They seem to exert their toxic effect at the level of the cell membrane. The precise function, in plants, of these proteins is not known.

CC -!- SIMILARITY: Belongs to the plant thionin family.

DR PIR; A01802; VFPD1T.

DR HSSP; P01544; 2PLH.

DR InterPro; IPR001010; Thionin.

DR Pfam; PF00321; plant thionins; 1.

DR PROSITE; PS00271; THIONIN; 1.

KW Plant defense; Thionin; Plant toxin.

FT DISULFID 3 40 PROBABLE.

FT DISULFID 4 32

FT DISULFID 16 26

FT MOD_RES 46 46 BLOCKED (BUT IS NOT AMIDATED).

SQ SEQUENCE 46 AA; 4881 MW; A02A08BA4E27D1A9 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 46;

Best Local Similarity 100.0%; Pred. No. 9.9;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 NTCRFGG 18

DB 14 NTCRFGG 20

RESULT 14

SUGB_CITFR STANDARD; PRT; 105 AA.

AC O69279;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE SugaE protein homolog.

GN SUGB.

OS Citrobacter freundii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Citrobacter.

OX NCBI_TaxID=546;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OS60;

RX MEDLINE=98344100; PubMed=9677290;

RA Bishop R.E., Leski B.K., Hodges R.S., Kay C.M., Weiner J.H.;

RT "The entericidin locus of Escherichia coli and its implications for programmed bacterial cell death.";

RL J. Mol. Biol. 280:583-596(1998).

CC -!- FUNCTION: COULD ACT AS A CHAPERONE (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -!- SIMILARITY: Belongs to the small multidrug resistance (SMR) protein family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: U21727; AAC46457.1; -
DR InterPro: IPR000390; Smr. Drug Res; 1.
DR Pfam: PF00893; Multi-Drug Res; 1.
KW Chaperone; Transmembrane; Transport.
FT TRANSMEM 1 21
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
SQ SEQUENCE 105 AA; 10897 MW; 62DA129ADA86E765 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 GESASPA 734
Db 79 GESASPA 85

RESULT 15
ACPS BRAJA
ID ACPS BRAJA STANDARD; PRT; 139 AA.
AC C69159;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase)
DE (4'-phosphopantetheinyl transferase acps).
CN ACPS OR BLL5063
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110SPC4;
RX MEDLINE=9908246; PubMed=9870699;
RA Baird A., Mueller P.;
RT "A second gene for type I signal peptidase in Bradyrhizobium
RT japonicum, sipF, is located near genes involved in RNA processing and
RT cell division."
RL Mol. Gen. Genet. 260:346-356(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
CC -!- FUNCTION: Transfers the 4'-phosphopantetheine moiety from coenzyme
CC A to a Ser of acyl-carrier protein (By similarity).
CC -!- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine
CC 3',5'-bisphosphate + holo-[acyl-carrier protein].
CC -!- COPACITOR: Magnesium (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the P-Pant transferase superfamily. Acps
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AF065159; AAD02937.1; -
DR EMBL: AP005953; BAC50328.1; -
DR HAMAP: MF 00101; -; 1.
DR InterPro: IPR008278; 4-PPT_transf.
DR InterPro: IPR002582; ACPS_trn.
DR InterPro: IPR004588; Pantethn_trn.
DR Pfam: PF01648; ACPS; 1.
DR ProDom: PD004282; ACPS; 1.
DR TIGRFAMs: TIGR00516; acps; 1.
DR TIGRFAMs: TIGR00556; pantethn_trn; 1.
KW Transferase; Lipid synthesis; Fatty acid biosynthesis; Magnesium;
KW Complete proteome.

FT METAL 8
FT METAL 61
SQ SEQUENCE 139 AA; 15193 MW; 2FD4A315845FC788 CRC64;
Query Match 0.9%; Score 7; DB 1; Length 139;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 522 ARIDVSI 528
Db 112 ARIDVSI 118

Search completed: August 25, 2004, 18:12:25
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:26:03 ; Search time 27 Seconds
(without alignments)

1496.535 Million cell updates/sec

Title: US-09-874-162A-8

Perfect score: 4078

Sequence: 1 MTGIAASFNTCRFGCG.....KALETDSVGSVKSKQKL 776

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|--------------------|
| 1 | 3413 | 83.7 | 739 | 1 JJAZ HUMAN | Q15022 homo sapien |
| 2 | 3332 | 81.7 | 741 | 1 JJAZ MOUSE | Q80u70 mus musculu |
| 3 | 1127.5 | 27.6 | 900 | 1 SUI2 DROME | Q9njg9 drosophila |
| 4 | 661 | 16.2 | 243 | 1 JAZ1 HUMAN | Q85vz6 homo sapien |
| 5 | 661 | 16.2 | 243 | 1 JAZ1 MOUSE | Q80zc5 mus musculu |
| 6 | 224 | 5.5 | 631 | 1 ENF2 ARATH | Q816y4 arabidopsis |
| 7 | 217 | 5.3 | 440 | 1 VRN2 ARATH | Q8wsb1 arabidopsis |
| 8 | 211 | 5.2 | 692 | 1 FIS2 ARATH | Q9znt9 arabidopsis |
| 9 | 176.5 | 4.3 | 1142 | 1 GIN4 YEAST | Q12263 saccharomyc |
| 10 | 161 | 3.9 | 1453 | 1 Y373 BOVIN | Q9tu23 bos taurus |
| 11 | 155.5 | 3.8 | 1755 | 1 YJ27 YEAST | P47098 saccharomyc |
| 12 | 152 | 3.7 | 1957 | 1 SPOF SCHPO | Q10411 schizosacch |
| 13 | 150 | 3.7 | 912 | 1 UBF3 YEAST | Q01477 saccharomyc |
| 14 | 147.5 | 3.6 | 1790 | 1 USO1 YEAST | P25386 saccharomyc |
| 15 | 147 | 3.6 | 2230 | 1 GOA4 HUMAN | Q13439 homo sapien |
| 16 | 146.5 | 3.6 | 1235 | 1 TRK1 YEAST | P12685 saccharomyc |
| 17 | 145 | 3.6 | 1744 | 1 TANA XENLA | Q01550 xenopus lae |
| 18 | 145 | 3.6 | 1953 | 1 BNI1 YEAST | P41832 saccharomyc |
| 19 | 143.5 | 3.5 | 698 | 1 YB6 YEAST | P38283 saccharomyc |
| 20 | 141.5 | 3.5 | 1328 | 1 YMD9 YEAST | Q03434 saccharomyc |
| 21 | 141 | 3.5 | 1875 | 1 MUP1 YEAST | Q02455 saccharomyc |
| 22 | 139.5 | 3.4 | 1085 | 1 YAF4 SCHPO | Q09863 schizosacch |
| 23 | 139.5 | 3.4 | 1833 | 1 ZEP2 HUMAN | P31629 homo sapien |
| 24 | 137 | 3.4 | 1969 | 1 Z292 HUMAN | O60281 homo sapien |
| 25 | 136 | 3.3 | 2867 | 1 RBP2 PLAVB | Q00799 plasmodium |
| 26 | 135.5 | 3.3 | 1544 | 1 ARHC HUMAN | Q9nzn5 homo sapien |
| 27 | 135 | 3.3 | 3210 | 1 CENF HUMAN | P49454 homo sapien |
| 28 | 135 | 3.3 | 5430 | 1 MACF HUMAN | Q9upn3 homo sapien |
| 29 | 135 | 3.3 | 5938 | 1 MAC4 HUMAN | Q96pk2 homo sapien |
| 30 | 134.5 | 3.3 | 1148 | 1 RFC1 HUMAN | P35251 homo sapien |
| 31 | 134.5 | 3.3 | 1972 | 1 BA2B HUMAN | Q9uif8 homo sapien |
| 32 | 133.5 | 3.3 | 1328 | 1 YWE4 YEAST | Q04711 saccharomyc |
| 33 | 132 | 3.2 | 2104 | 1 MYS3 SCHPO | Q14157 schizosacch |

ALIGNMENTS

RESULT 1

| ID | JJAZ HUMAN | STANDARD; | PRT; | 739 AA. |
|----|--|-----------|------|---------|
| AC | Q15022; Q96BD9; | | | |
| DT | 10-OCT-2003 (Rel. 42, Created) | | | |
| DT | 10-OCT-2003 (Rel. 42, Last sequence update) | | | |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update) | | | |
| DE | Polycomb protein SUZ12 (Suppressor of zeste12 protein homolog) | | | |
| DE | (joined to JAZF1 protein) (Chromatin precipitated E2f target 9 protein). | | | |
| GN | SUZ12 OR JJAZ1 OR CHET9 OR KIAA0160. | | | |
| OS | Homo sapiens (Human). | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | TISSUE=Bone marrow; | | | |
| RX | MEDLINE=96127530; PubMed=850280; | | | |
| RA | Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.; | | | |
| RT | "Prediction of the coding sequences of unidentified human genes. IV. | | | |
| RT | The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by | | | |
| RT | analysis of cDNA clones from human cell line KG-1." | | | |
| RL | DNA Res. 2:167-174(1995). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | TISSUE=Uterus; | | | |
| RX | MEDLINE=2238257; PubMed=12477932; | | | |
| RA | Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., | | | |
| RA | Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., | | | |
| RA | Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., | | | |
| RA | Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., | | | |
| RA | Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., | | | |
| RA | Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., | | | |
| RA | Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., | | | |
| RA | Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., | | | |
| RA | Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., | | | |
| RA | Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., | | | |
| RA | Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., | | | |
| RA | Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., | | | |
| RA | Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., | | | |
| RA | Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., | | | |
| RA | Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., | | | |
| RA | Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smailus D.E., | | | |
| RA | Schmurch A., Schein J.E., Jones S.J.M., Marra M.A.; | | | |
| RT | "Generation and initial analysis of more than 15,000 full-length human | | | |
| RT | and mouse cDNA sequences"; | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002). | | | |
| RN | [3] | | | |
| RP | SEQUENCE OF 102-113; 343-360; 474-483 AND 509-516, IDENTIFICATION IN A | | | |
| RP | PRC2 COMPLEX WITH EED; EZH2; RBBP4 AND RBBP7, AND METHYLTRANSFERASE | | | |
| RP | ACTIVITY OF THE COMPLEX. | | | |
| RX | MEDLINE=22300550; PubMed=12351676; | | | |
| RA | Cao R., Wang L., Wang H., Xia L., Erdjument-Bromage H., Tempst P., | | | |
| RA | Jones R.S., Zhang Y.; | | | |
| RT | "Role of histone H3 lysine 27 methylation in Polycomb-group | | | |

Q04670 saccharomyc
P54697 dictyosteli
Q96t58 homo sapien
Q12955 homo sapien
P04932 plasmodium
P04933 plasmodium
Q03661 saccharomyc
P15619 drosophila
P47100 saccharomyc
Q92056 mus musculu
P46100 homo sapien
Q7yqm4 pan troglod

silencing.";
 Science 298:1039-1043 (2002).
 [4]
 RX DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH JAZF1.
 MEDLINE=21265470; PubMed=11371647;
 RA Koontz J.I., Soreng A.L., Nucci M., Kuo P.C., Pauwels P.,
 RA van Den Berghe H., Cin P.D., Fletcher J.A., Sklar J.;
 RT "Frequent fusion of the JAZF1 and JAZ2 genes in endometrial stromal
 tumors.";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353 (2001).
 [5]
 RX MEDLINE=214448978; PubMed=11564866;
 RA Weinmann A.S., Bartley S.M., Zhang T., Zhang M.Q., Farnham P.J.;
 RT "Use of chromatin immunoprecipitation to clone novel E2F target
 promoters.";
 RT Mol. Cell. Biol. 21:6820-6832 (2001).
 [6]
 RP IDENTIFICATION IN A PRC2 COMPLEX WITH EED; EZH2; RBBP4 AND RBBP7, AND
 METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
 RX MEDLINE=22323099; PubMed=12435631;
 RA Kuzmichev A., Nishio K., Erdjument-Bromage H., Tempst P.,
 RA Reinberg D.;
 RT "Histone methyltransferase activity associated with a human
 multiprotein complex containing the Enhancer of Zeste protein.";
 RL Genes Dev. 16:2893-2905 (2002).
 CC -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by
 forming multiprotein complexes, which are required to maintain the
 transcriptionally repressive state of homeotic genes throughout
 development. PcG proteins are not required to initiate repression,
 but to maintain it during later stages of development. They
 probably act via the methylation of histones, rendering chromatin
 heritably changed in its expressibility. Component of the PRC2
 complex, which methylates Lys-9 and Lys-27 residues of histone H3.
 CC -!- SUBUNIT: Component of the PRC2 complex, composed of EED, EZH2,
 SUZ12/JAZ2, RBBP4 and RBBP7. This complex is distinct from the
 PRC1 complex, which contains many other PcG proteins like BMI1,
 CBX4, CBX8, PHC2, PHC3, SCMH1, RING1 and RNF2.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- INDUCTION: Induced by E2F1 transcription factor.
 CC -!- DISEASE: Endometrial stromal tumors can be caused by a chromosomal
 rearrangement involving JAZF1 and SUZ12. A t(7;12)(p15;q21)
 chromosomal translocation involving JAZF1 and SUZ12 generates the
 JAZF1-SUZ12 oncogene consisting of the N-terminus part of JAZF1
 and the C-terminus part of SUZ12. The translocation is frequently
 found in all cases of endometrial stromal tumors, except in
 endometrial stromal sarcomas, where it is rarer.
 CC -!- SIMILARITY: Belongs to the VEFs (VRN2-EMF2-FIS2-SU(Z)12) family.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D63881; BAA09931.1; ALT INIT.
 DR EMBL; BC015704; AAH15704.1; -;
 DR MIM; 606245; -;
 DR InterPro; IPR007087; Znf_C2H2.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
 KW Transcription regulation; Repressor; Chromatin regulator;
 KW Nuclear protein; Metal-binding; Zinc; Zinc-finger;
 KW Chromosomal translocation; Proto-oncogene.
 FT ZN FING 448 471 C2H2-TYPE.
 FT DOMAIN 563 639 VEFs-BOX.
 FT DOMAIN 7 50 GLY-RICH.
 FT DOMAIN 51 59 POLY-SER.
 FT DOMAIN 60 67 POLY-ALA.
 FT SITE 93 94 BREAKPOINT FOR TRANSLOCATION TO FORM

FT CONFLICT 216 216 JAZF1-SUZ12 ONCOGENE.
 SQ SEQUENCE 739 AA; 83053 MW; A8830BBD031DF656 CRC64;
 I -> N (IN REF. 2).
 Query Match 83.7%; Score 3413; DB 1; Length 739;
 Best Local Similarity 99.8%; Pred. No. 5.6e-199;
 Matches 647; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 129 BTOYRFLRTRNLAPIPLHRLTYMSHRSRTNKKTKVDDMLSKVEMKGEQESH 188
 Db :|||||
 QY 92 KPTQYRFLRTRNLAPIPLHRLTYMSHRSRTNKKTKVDDMLSKVEMKGEQESH 151
 QY 189 SLSAHLQLTFTGFFHNKDKPSNSEQNSVTLVLLVKVCHKKKDVSCIPRQVPTGKK 248
 Db :|||||
 QY 152 SLSAHLQLTFTGFFHNKDKPSNSEQNSVTLVLLVKVCHKKKDVSCIPRQVPTGKK 211
 QY 249 QVPLIPDLNQTGKGNFPSPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRRENGMNGETN 308
 Db :|||||
 QY 212 QVPLIPDLNQTGKGNFPSPSLAVSSNEFEPSNSHMVKSYSLLFRVTRPGRRENGMNGETN 271
 QY 309 ENIDVNEELPARKKRNRDEGKTFVAQMTVPDKNRRLQLDGEYEVAMQEMEECPISKKR 368
 Db :|||||
 QY 272 ENIDVNEELPARKKRNRDEGKTFVAQMTVPDKNRRLQLDGEYEVAMQEMEECPISKKR 331
 QY 369 ATWETILDGKRLPPPTETFSQGTPLQFTLRWTGTTNDKSTAPIAKPLATRNSESILHQENKP 428
 Db :|||||
 QY 332 ATWETILDGKRLPPPTETFSQGTPLQFTLRWTGTTNDKSTAPIAKPLATRNSESILHQENKP 391
 QY 429 GSVKPTQITAVKESITTDLQTRKXDTNENKQKRIFYQFLYNNNTTQCTEARDLHCP 488
 Db :|||||
 QY 392 GSVKPTQITAVKESITTDLQTRKXDTNENKQKRIFYQFLYNNNTTQCTEARDLHCP 451
 QY 489 WCTLNCRKLYSLKHLKLSHRSFIENYVYHPKGRIDVSIENECYDGSVAGNPQDIHQPG 548
 Db :|||||
 QY 452 WCTLNCRKLYSLKHLKLSHRSFIENYVYHPKGRIDVSIENECYDGSVAGNPQDIHQPG 511
 QY 549 FAFSRNGPVKPTPIHILVCRPKRTKASMSPLESEDEGEVQOORTYSSGHNLRFHSDTC 608
 Db :|||||
 QY 512 FAFSRNGPVKPTPIHILVCRPKRTKASMSPLESEDEGEVQOORTYSSGHNLRFHSDTC 571
 QY 609 LPLRQCEMEVDESEKDPWLREKTIITQIEFSDVNEGEKVMKLNHLVHKHGFADNQ 668
 Db :|||||
 QY 572 LPLRQCEMEVDESEKDPWLREKTIITQIEFSDVNEGEKVMKLNHLVHKHGFADNQ 631
 QY 669 MNHACMLFVENVYQKIIKKNLCRNFMHLVSHMDFNLISIMSIDKAVTKLREMOKLEKG 728
 Db :|||||
 QY 632 MNHACMLFVENVYQKIIKKNLCRNFMHLVSHMDFNLISIMSIDKAVTKLREMOKLEKG 691
 QY 729 ESASPAHEITEFQNGTANGSEINSKEALETDSVSGVSKQSKQKL 776
 Db :|||||
 QY 692 ESASPAHEITEFQNGTANGSEINSKEALETDSVSGVSKQSKQKL 739

RESULT 2

JJAZ_MOUSE
 ID JJAZ_MOUSE STANDARD; PRT; 741 AA.
 AC Q80U70; Q80Y10; Q99L07;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Polycomb protein Suz12 (Suppressor of zeste 12 protein homolog).
 GN SUZ12 OR KIAA0160 OR D11ERTD530E.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID:10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
 RA Nakajima D., Nagase T., Ohara O., Koga H.;
 RA "Prediction of the coding sequences of mouse homologues of KIAA gene:

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA April J.F., Abayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dalke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.S., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriter S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glöckle A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hartis N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kratt C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stimpson M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RX MEDLINE=2426069; PubMed=12537562;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kinkler J.S., Millburn G.H., Prochownik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC STRAIN=Berkeley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Kronmiller B., Paclebb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A *Drosophila* full-length cDNA resource";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 RN [5]
 RP IDENTIFICATION IN A ESC/E(2) COMPLEX WITH E(2); CAF1 AND ESC,
 RP AND METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
 RX MEDLINE=22296673; PubMed=12408863;
 RA Czermin B., Melfi R., McCabe D., Seitz V., Imhof A., Pirrotta V.,
 RT "Drosophila enhancer of Zeste/ESC complexes have a histone H3
 RT methyltransferase activity that marks chromosomal Polycomb sites";
 RL Cell 111:185-196(2002).
 RN [6]
 RP IDENTIFICATION IN A ESC/E(2) COMPLEX WITH E(2); CAF1 AND ESC, AND
 RP METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
 RX MEDLINE=22296674; PubMed=12408864;
 RA Mueller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A.,

RA Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A.;
 RT "Histone methyltransferase activity of a *Drosophila* Polycomb group
 RT repressor complex";
 RL Cell 111:197-208(2002).
 CC -1- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by
 CC forming multiprotein complexes, which are required to maintain the
 CC transcriptionally repressive state of homeotic genes throughout
 CC development. PcG proteins are not required to initiate repression,
 CC but to maintain it during later stages of development. They
 CC probably act via the methylation of histones, rendering chromatin
 CC heritably changed in its expressibility. Component of the Esc/E(2)
 CC complex, which methylates Lys-9 and Lys-27 residues of histone H3.
 CC Despite the presence of a zinc-finger, it does not bind directly
 CC to DNA, the Esc/E(2) complex being probably recruited to DNA by
 CC Pho. The Esc/E(2) complex is necessary but not sufficient to
 CC recruit a functional PcG repressive complex that represses target
 CC genes, suggesting that the recruitment of the distinct PRC1
 CC complex is also required to allow a subsequent repression.
 CC -1- SUBUNIT: Component of the Esc/E(2) complex, composed of Esc, E(z),
 CC Su(z)12, Rpd3, Caf1 and probably Pho. This complex is distinct
 CC from the PRC1 complex, which contains many other PcG proteins like
 CC Pc, Ph, Psc, Su(z)2. The two complexes however cooperate and
 CC interact together during the first 3 hours of development to
 CC establish PcG silencing.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=B;
 CC IsoId=Q9NUG9-1; SequencesDisplayed;
 CC Name=2; Synonyms=A;
 CC IsoId=Q9NUG9-2; Sequence=VSP_007033, VSP_007034;
 CC Note=No experimental confirmation available;
 CC -1- SIMILARITY: Belongs to the VSPs (VRN2-EMF2-FIS2-SU(Z)12) family.
 CC -1- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF149047; AAF73149.1; -;
 CC EMBL: AE003515; AAF49094.2; -;
 CC EMBL: AE003515; RAN11641.1; -;
 CC EMBL: AY069809; RAL39954.1; -;
 CC FlyBase: FBgn0020887; Su(z)12.
 CC GO: GO:0000790; C:nuclear chromatin; NAS.
 CC GO: GO:0003677; F:DNA binding; NAS.
 CC GO: GO:0016458; P:gene silencing; NAS.
 CC GO: GO:0016481; P:negative regulation of transcription; NAS.
 CC InterPro: IPR007087; Znf_C2H2;
 CC SMART: SM00355; Znf_C2H2; 1;
 CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1;
 CC PROSITE: PS00157; ZINC_FINGER_C2H2_2; FALSE NEG.
 CC Transcription regulation; Repressor; Developmental protein;
 CC Nuclear protein; Metal-binding; Zinc; Zinc-finger;
 CC Alternative splicing.
 CC ZN_RING 411 434 C2H2-TYPE.
 CC DOMAIN 527 603 VEF5-BOX.
 CC DOMAIN 355 366 ASN-RICH.
 CC DOMAIN 699 879 SER-RICH.
 CC VARSPLIC 806 855
 CC VARSPLIC 856 900 Missing (in isoform 2).
 CC MUTAGEN 274 274 G->D: IN SU(Z)12-2; INDUCES LARVAL
 CC SEQUENCE 900 AA; 100104 MW; 53BA0D83C49BC92F CRC64;

```
Query Match      27.6%; Score 1127.5; DB 1; Length 900;
Best Local Similarity 32.9%; Pred. No. 9.3e-61;
Matches 273; Conservative 137; Mismatches 240; Indels 179; Gaps 26;

QY 65 MTDAAAREOESLKKIQPKSLTLSSVSRGNV-----STPPRHSSGSLTPPVTPITPSS 120
DB 1 MAPAKKREKDS-----NPDGSAANGIIGLTHGAPDASNAGSVPTPAEQOVKLN 49
QY 121 SFSSSTP-----TEPTQIYRLRNLAPIFLHRTLYTMSHRNSRNIRKTKPKVDDM 174
DB 50 GHOQOEQLFQAFKPTQIYRLRNLAPIFLHRTLYTMSHRNSRNIRKTKPKVDDM 174
QY 175 LSKVKMKGEQESHLSA-HLQITFTGFHKKDKSPSENQ-----NSVTLLE 222
DB 110 LESI-----TQSEAVSQNYLHVYDSLHEKLPARLDNESHGDLLOEQLLCEAGESVSVE 164
QY 223 VLLVVKVCHKRDVSCPIRQVPTGKQVPLIPDLNQTGKGNPSPVLAASNFEP-SNSHM 281
DB 165 TLYKITRSKRDSTLDFOELLKSCQIVYNP---KDRVGEHATISPIQWTRPMGEQHT 221
QY 282 VKSYLLFRVTRPGRRFNGMNGENINIDVNEELPARKKENREDGKTFVAQMTVFDK 341
DB 222 L-YKLLFRK-----VLSPTCN--DENATPP-NKSRPN-EMFGSEILILEYK 266
QY 342 NRRLOLDGEYEVAMQEMECPI---SKRATWETILDGKRLP---PFETFSQGETLOFT 395
DB 267 SSGF-ITEGEYEAQLPLNSTSIKFSFKCTWETMPD-SYIPLSLTYDVVYQSPMLKXF 324
QY 396 LWTGETNDKSTAPIAKPLATR-----NSESLOEN-----KGSVKPTQTI 437
DB 325 LTL---SNQLEPMISAPLOQVYQVHLDVAEMNYNNNNNNNNCSGLKNGSGGNTV 381
QY 438 AVKESITLDTQKREKDTFNENRQKRLIPYQPLNNTRQOFEARDDLHCPCWCTLNCRKL 497
DB 382 C-----KTPP-----EHIQIVNFWYNNTRQOFEYQELNCPWGLDCLRL 423
QY 498 YSLKHLKLCRGRPFVYVHPKGRDIVSINECVDSYAGNPQDIHQPGAFPSRN-GP 556
DB 424 YALLKHLKLCRGRPFVYVHPKGRDIVSINECVDSYAGNPQDIHQPGAFPSRN-GP 556
QY 557 VKRTPITHLVCRPKTKASMEERLESDEGEVQOQRTSSGHNRLYFHSDFCLPLRPOEM 616
DB 484 VRTSVTSLMVCRRPKQKCTLDEFLDEDELSNORSYITGHNRLYFHSDFCLPLRPOEM 543
QY 617 EYVDEKDPPEWLREKTIQIEEFSDVNEGEKVMKLNLMVHMGHFIADQNMHACMLF 676
DB 544 DIDSGESDPLRQKTIQIMIDEFSDVNEGEKVMKLNLMVHMGHFIADQNMHACMLF 603
QY 677 VENYQKLIKKNLCLNFMHLVSMHDFNLISIMSDKAVTKLREMQQKLEKGES----- 730
DB 604 LDKAGTEIVRKNLYRNFLILMCSLFDYGLIAETVYKTVQKLGLLSKYAAGQELMQRQR 663
QY 731 -----ASPAEEITE----- 740
DB 664 BEQLXVLDVGMHKKQEDPKTLKSPQKAPPADQASTSSATSGSGSSSSNQPKKMPA 723
QY 741 -----FONGTANGSEINSKEKALETDSVSGSKQKQKL 776
DB 724 HLKRGSAASSPGVQSGKTENGNGNS-----SSNSNKNVAKSADQPL 767
```

RESULT 4

```
JAZ1_HUMAN
ID JAZ1_HUMAN STANDARD; PRT; 243 AA.
AC Q86VZ6; Q8N3L7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Juxtaposed with another zinc finger protein 1.
GN JAZF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
TISSUE=Brain, and Testis;
MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klauser R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.A., Lequellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
FAhey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Kzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).
TISSUE=Amalgam;
Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;
Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
[3]
DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH SUZ12.
MEDLINE=21265470; PubMed=11371647;
RA Koontz J.I., Sorong A.L., Nucci M., Kuo F.C., Pauwels P.,
van Den Berghe H., Cin P.D., Fletcher J.A., Sklar J.;
"Frequent fusion of the JAZF1 and JAZF2 genes in endometrial stromal
tumors."
Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353(2001).
-!- FUNCTION: Potential transcription factor.
-!- SUBCELLULAR LOCATION: Nuclear protein (Potential).
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
IsoId=Q86VZ6-1; Sequence=Displayed;
Name=2;
IsoId=Q86VZ6-2; Sequence=VSP_007755;
Note=Sequence incomplete. No experimental confirmation
available.
-!- DISEASE: Endometrial stromal tumors can be caused by a chromosomal
rearrangement involving JAZF1 and SUZ12. A t(7;17)(p15;q21)
chromosomal translocation involving JAZF1 and SUZ12 generates the
JAZF1-SUZ12 oncogene consisting of the N-terminus part of JAZF1
and the C-terminus part of SUZ12. The translocation is frequently
found in all cases of endometrial stromal tumors, except in
endometrial stromal sarcomas, where it is rarer.
-!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; BC042441; AAH42441.1; --
EMBL; AL834234; CAD38912.1; --
MIM; 606246; --
InterPro; IPR007087; Znf_C2H2.
Pfam; PF00096; zf-C2H2; 3.
SMART; SM00355; Znf_C2H2; 3.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
```

KW Transcription regulation; Nuclear protein; Metal-binding; Zinc;
 KW Zinc-finger; Chromosomal translocation; Proto-oncogene;
 KW Alternative splicing;
 FT ZN_FING 12 37 C2H2-TYPE 1.
 FT ZN_FING 173 198 C2H2-TYPE 2.
 FT ZN_FING 208 230 C2H2-TYPE 3 (DEGENERATE).
 FT SITE 129 130 BREAKPOINT FOR TRANSLLOCATION TO FORM
 FT VARSPLIC 1 38 JAZF1-SUZ12 ONCOGENE
 FT MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNH1 ->
 FT YFVLSGLIYSDAKSKKRNKCNVIT (in isoform 2).
 FT CONFLICT 133 133 /FTID=VSP 007755.
 FT D -> G (IN REF. 1).
 FT SEQUENCE 243 AA; 27079 MW; 878A6E3D82C0588 CRC64;
 Query Match 16.2%; Score 661; DB 1; Length 243;
 Best Local Similarity 100.0%; Pred. No. 2.7e-33;
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTVALSY 60
 DB 1 MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTVALSY 60
 QY 61 INRFMTDAARBOESLKKIOPKSLTSSVSRSRGVSTPPRHSSGSLTPVTPPTSS 120
 DB 61 INRFMTDAARBOESLKKIOPKSLTSSVSRSRGVSTPPRHSSGSLTPVTPPTSS 120
 QY 121 SPFSSTPT 128
 DB 121 SPFSSTPT 128
 RESULT 5
 ID JAZ1 MOUSE STANDARD; PRT; 243 AA.
 AC Q80Z05; Q8BL08; Q8BNW4;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Juxtaposed with another zinc finger protein 1.
 GN JAZF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/6J; TIGSUS=Brain cortex, and Pancreas;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yananaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bragdi D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Giesi C., Godzik A., Gough J.,
 RA Grimond S., Gusninch S., Hirokawa N., Jackson I.J., Jarvis B.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Resole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Sempé C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Vetardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hironaka-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imorani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Havashizaki Y.,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 RP [2]
 RC SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Shat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hsieh F.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- FUNCTION: Potential transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear protein (Potential).
 CC -!- ALTERNATIVE PRODUCTS:
 CC -!- Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoID=Q80ZQ5-1; Sequence=Displayed;
 CC Name=2;
 CC IsoID=Q80ZQ5-2; Sequence=VSP 007756;
 CC Note=No experimental confirmation available;
 CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AK043767; BAC31649.1; -;
 CC EMBL; AK050474; BAC34275.1; -;
 CC EMBL; BC048577; AAH48577.1; -;
 CC InterPro; IPR007087; Znf_C2H2.
 CC Pfam; PF00096; zf-C2H2; 3.
 CC SMART; SM00355; Znf_C2H2; 3.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; FALSE NEG.
 CC Transcription regulation; Nuclear protein; Metal-binding; Zinc;
 CC Zinc-finger; Alternative splicing.
 KW ZN_FING 12 37 C2H2-TYPE 1.
 FT ZN_FING 173 198 C2H2-TYPE 2.
 FT ZN_FING 208 230 C2H2-TYPE 3 (DEGENERATE).
 FT VARSPLIC 64 64 Missing (in isoform 2).
 FT /FTID=VSP 007756.
 FT CONFLICT 211 211 R -> L (IN REF. 1; BAC34275).
 FT SEQUENCE 243 AA; 27097 MW; 878A6E3D82C0348 CRC64;
 SQ
 Query Match 16.2%; Score 661; DB 1; Length 243;
 Best Local Similarity 100.0%; Pred. No. 2.7e-33;
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTVALSY 60
 DB 1 MTGIAAASFNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTVALSY 60

CC forming repressor, after a cold treatment. PcG proteins act by
 CC forming multiprotein complexes, which are required to maintain the
 CC transcriptionally repressive state of homeotic genes throughout
 CC development. PcG proteins are not required to initiate repression,
 CC but to maintain it during later stages of development. They
 CC probably act via the methylation of histones, rendering chromatin
 CC heritably changed in its expressibility.
 CC -!- SUBUNIT: Probable component of a PcG complex. In plants, PcG
 CC complexes are probably composed of a member of the EZ family (CLF
 CC or MEA), FIE, and a member of the VEFs family (FIS2, VRN2 or EBF2)
 CC (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9W5B1-1; Sequence=Displayed;
 CC Name=2; Synonyms=VRN2;
 CC IsoId=Q9W5B1-2; Sequence=VSP_007457, VSP_007458;
 CC TISSUE SPECIFICITY: Weakly expressed. Expressed both during, and
 CC in the absence of vernalization.
 CC -!- SIMILARITY: Belongs to the VEFs (VRN2-EMF2-FIS2-SU(2)12) family.
 CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to
 CC erroneous gene model prediction.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF284500; AAL32135.1; -;
 DR EMBL; AF284501; AAL32136.1; -;
 DR EMBL; Z97342; CAB10457.1; ALT_SEQ.
 DR EMBL; AL161545; CAB80955.1; ALT_SEQ.
 DR EMBL; AY034902; AAK59409.1; -;
 DR EMBL; AY063047; AAL34221.1; -;
 DR InterPro; IPR007087; Znf_C2H2.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; FALSE NEG.
 KW Transcription regulation; Repressor; Nuclear protein; Zinc;
 KW Zinc-finger; Metal-binding; Alternative splicing.
 FT ZN_FING 86 111
 FT DOMAIN 156 163 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 267 345 VEFs-BOX.
 FT DOMAIN 398 402 POLY-ASN.
 FT DOMAIN 411 416 POLY-ASN.
 FT VARSPIC 102 107 GLOFHL -> VGNYYN (in isoform 2).
 FT VARSPIC 108 440 Missing (in isoform 2).
 FT CONFLICT 81 81 R -> K (IN REF. 1).
 FT CONFLICT 121 121 L -> S (IN REF. 1).
 FT CONFLICT 185 185 A -> T (IN REF. 1).
 FT CONFLICT 193 193 A -> T (IN REF. 1).
 FT CONFLICT 229 229 T -> S (IN REF. 1).
 FT CONFLICT 344 344 V -> A (IN REF. 1).
 FT CONFLICT 355 355 C -> R (IN REF. 1).
 FT CONFLICT 394 398 TSVTN -> SSDTTT (IN REF. 1).
 FT CONFLICT 406 406 H -> R (IN REF. 1).
 FT CONFLICT 425 425 K -> N (IN REF. 1).
 FT CONFLICT 440 440 K -> KVTK (IN REF. 1).
 CC SEQUENCE 440 AA; 50623 MW; 0E0AB2C5517BE6F CRC64;

Query Match 5.3%; Score 217; DB 1; Length 440;
 Best Local Similarity 21.2%; Pred. No. 3.9e-06;
 Matches 80; Conservative 62; Mismatches 140; Indels 96; Gaps 11;
 450 RKEKTPNRCOKLRIFFLYNNTRQTEARDLHCPWCLNCRKLYSLKHLKCHS 509
 58 RRRK-----SRSTGMVFNKYDCNNTLQRTVEEDCSFPCMLCGSPKGLQFLNNSHD 112

QY 510 RFIF-----NYYVHPKGARDV-----SI 528
 DB 113 LFEEFKLLEBYQTVNVSVKLNISFIIEEGSDDDKFFPFLSKPKRRKRRGGNNTRRL 172
 QY 529 NECY-----DGSYA-GNPQDIHRQPGFAPSRNGPVKRTDPIH-- 564
 DB 173 KVCPLPLDSPSLANGTENGIALNDGNRGLGYPEATELAGQFEMTSNIP---PAIAHSSL 229
 QY 565 -----ILVCPKPKTKASMEFSEDEGEVQQRYSCHNRLYFHSDTCLPLRQEM--E 617
 DB 230 DAGAKVILITTEAVVPATKTRKLKSAERSEA---RSHLLQKQRFVHSRVQFMALEQVMSD 286
 QY 618 VDSDEKDPWLREKTTITQIEFSDVNVGEKVKLWNLHVMKHGFIADQMAHACMLFV 677
 DB 287 RDSDEVDVDDVADFEQMLDDFDVNVKDEKQFHLWNSFVKQKRVIADGHSWACEVFS 346
 QY 678 ENYQOKI-IKKNLCRNPMHLVSMHDPNLISISIDKAVTKLREMQOKLKGSSASAPNE 736
 DB 347 RFYEKELHCYSSFLWCWRLFLIKLWNHGLVDSATINNCNTILENCR-----NT 394
 QY 737 EITEEQNGTANGPSEINS 754
 DB 395 SVTNNNNSVDHPDSNT 412
 RESULT 8
 FIS2 ARATH
 ID FIS2 ARATH STANDARD; PRT; 692 AA.
 AC Q9ZNT9; Q9ZQP0;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Polycomb group protein FERTILIZATION-INDEPENDENT SEED 2.
 GN FIS2 OR AT2G35670 OR t20F21.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A., MUTANT FIS2-4, AND VARIANTS.
 RP STRAIN=cv. Landsberg erecta; TISSUE=Silique;
 RX MEDLINE=99093530; PubMed=9874812;
 RA Luo M., Bilodeau P., Koltunow A., Dennis E.S., Peacock W.J.,
 RA Chaudhury A.;
 RT "Genes controlling fertilization-independent seed development in
 RL Arabidopsis thaliana";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:296-301(1999).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Rensing C.M., Koo H.L.,
 RA Moffat K.S., Cronin L.A., Shen W., Pai G., Van Aken S., Umayam L.,
 RA Tallon L.J., Gill J.B., Adams M.D., Carrera A.J., Creasy T.H.,
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RL thaliana";
 RL Nature 402:761-768(1999).
 [3]
 RN SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RP MEDLINE=20442428; PubMed=10962025;
 RX Luo M., Bilodeau P., Dennis E.S., Peacock W.J., Chaudhury A.;
 RT "Expression and parent-of-origin effects for FIS2, MEA, and FIE in the
 RT endosperm and embryo of developing Arabidopsis seeds";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:10637-10642(2000).
 CC -!- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by
 CC forming multiprotein complexes, which are required to maintain the

transcriptionally repressive state of homeotic genes throughout development. PCG proteins are not required to initiate repression, but to maintain it during later stages of development. They probably act via the methylation of histones, rendering chromatin heritably changed in its expressibility. Required to prevent the proliferation of the central cell by repressing unknown target genes before fertilization.

CC -!- SUBUNIT: Probably indirectly associated with FIE and/or MEA. In plants, PCG complexes are probably composed of a member of the EZ family (CLF or MEA), FIE, and a member of the VEFs family (FIS2, VRN2 or EMF2) (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- TISSUE SPECIFICITY: Weakly expressed. Expressed in late silicles.

CC -!- DEVELOPMENTAL STAGE: Expressed maternally and zygotically. Expressed in the central cell before fertilization, and in the endosperm after fertilization, then decreases before the time of endosperm cellularization but continues in the chalazal cyst.

CC -!- POLYMORPHISM: In cv. Columbia, the sequence differs from that shown due to a deletion in the genomic sequence that covers the Leu-443 to Ile-502 region.

CC -!- SIMILARITY: Belongs to the VEFs (VRN2-EMF2-FIS2-SU(Z)12) family.

CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF096095; AAD09104.1; -;
 DR EMBL; AF096096; AAD09105.1; -;
 DR EMBL; AF096096; AAD15448.2; ALT_SEQ.
 DR InterPro; IPR007087; Znf_C2H2.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; FALSE NEG.
 KW Transcription regulation; Repressor; Nuclear protein;
 KW Developmental protein; Repeat; Zinc-finger; Metal-binding;
 KW Polymorphism.
 FT ZN_FING 11 34
 FT DOMAIN 122 502
 FT REPEAT 122 143
 FT REPEAT 161 183
 FT REPEAT 184 206
 FT REPEAT 207 228
 FT REPEAT 229 250
 FT REPEAT 259 280
 FT REPEAT 299 320
 FT REPEAT 338 360
 FT REPEAT 361 381
 FT REPEAT 399 421
 FT REPEAT 422 442
 FT REPEAT 459 480
 FT REPEAT 481 502
 FT DOMAIN 144 519
 FT REPEAT 144 160
 FT REPEAT 251 267
 FT REPEAT 282 298
 FT REPEAT 321 337
 FT REPEAT 382 398
 FT REPEAT 443 458
 FT REPEAT 503 519
 FT DOMAIN 543 621
 FT VARIANT 269 269
 FT VARIANT 442 501
 FT VARIANT 554 554
 FT MUTAGEN 637 692
 FT SEQUENCE 692 AA; 77666 MW; 2A3F61B961676497 CRC64;
 Query Match 5.2%; Score 211; DB 1; Length 692;

Best Local Similarity 17.8%; Pred. No. 1.6e-05;
 Matches 149; Conservative 122; Mismatches 286; Indels 278; Gaps 30;
 QY 19 CGLPFTLADLIEHEDNH-----IDTPRV-----LKKQELQQPTY 55
 Db 20 CGGH-----EGLQLHLSHDAFKFERYAEKDHGPEVDVSKSDTIKFGVLKDDVGNPQL 75
 QY 56 VALSYINRFTDAAREQESLKKIKQ-----PKLSLTSSSVSRGNVSPR-H 103
 Db 76 SPITFCCKNRNRRORDDSNVKNLVLMELDDLPRGTENDSTHVDNVSFPRAH 135
 QY 104 SSGSLTPPTPTPTSSSPRSSTPTPTQIVRFLRLIAPIFLRLTLYMHSNSTN 163
 Db 136 SSEKISDILT--TTQALAESEPKVP-----HVNDGNVSPRAHSS---AEKNSTH 184
 QY 164 IRRKFKVDDMLSKVEKMGQESLSAHLQLTITGFHKNDKSPNSN----- 214
 Db 185 VN-----DD-----DDVSSPPRAHSL-----EKNESTHVEDNISSPFKAHS 221
 QY 215 -EONSVTLEVLVVKVCHKKRDKVSCPIR-----QVPTGKQVPLIPDLNQTGPN-- 263
 Db 222 SKKNST-----HWNDEDVSFPPTRSSKETSDILTTQPAIVFSEPKVRRGSR 272
 QY 264 -----FPLAVSS-----NEPEPSNMVKSYSLLFVRTRPG----- 295
 Db 273 KQIYAKRYKARETOPAIAESESPEKVLHVNDENVSSPPEAHSLEKASDILTTQPAIAESS 332
 QY 296 -----RRFENGINGETNENIDVNEELPARRKRNDEGKTFVAQMTVFD 340
 Db 333 EPKVPVHNDENVSTPRAHSSKKNKSTRKVD--NVSPSPKTRSSKSTSDILTTQPTIAE 391
 QY 341 KNRRLQLLDGEYAVQMECEPIISKRAWTETILDGKRLPPFTFSOGTQLQFLRWG 400
 Db 392 SS-----EPKVRHVDNDVSS-----TPRAHS 413
 QY 401 ETNDKSTAPIAKPLATRNSESLSHQENKPGSVKPTOTIAKESLTTDLQTRKEKDTNENR 460
 Db 414 SKKNKST-----RKNDNDNIPSPKTRSSKSTNLTTR---TOPAIAESEPKVPVHND 462
 QY 461 QKLRIFYQFLYNNTRTQQTAEADD--LHCPWCTLCNCRKLYSLKLKLCCHSRFIENVYHP 519
 Db 463 DKVSSPTR-AHSSKKNKSTHKDDNASLPPKTRSSKSTSDIL----- 503
 QY 520 KGARDVSNICYDGSYAGNPQDIHQFGFAPSRNGPKRPIITHILVCRPKRYKASME 579
 Db 504 -----ATTQPAKAEF-----SEPKVTRVSRK 525
 QY 580 FLESDGEVEQRTYSSGHNRLYFHSHTCLPLRPOEM--EVDSEDEKDPWELREKTIQI 637
 Db 526 ELHAERCEAKRLERLK--GRQFYHSQTMQPMTPFQVMSNEDSENEDDYALDLSERLRL 582
 QY 638 EEFSDVNEGEKEVMKLVNLMHVMKHGFIADNQNMHACMLFVE-----NYGQKII 685
 Db 583 ERLVGVSKEEKRYMYLWNIIFVRKQVRIADGHVPWACEEFAKLHKEEMKNSSFDWWRMF 642
 QY 686 KKNLCNFMHLVSMVHDNLSIMSIDKAVTKLRWQOKLEKESASPAHEITE 740
 Db 643 RIKLWNGLICAKTFHKCTTILLNSDEA-----GQFTSGSANNANQOSME 689

RESULT 9
 GIN4 YEAST
 ID_GIN4 YEAST STANDARD; PRT; 1142 AA.
 AC Q12283;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Serine/threonine-protein kinase GIN4 (EC 2.7.1.-).
 GN GIN4 OR YDR507C OR D9719.13.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;

[1] SEQUENCE FROM N.A.
 RP MEDLINE=99030835; PubMed=9813093;
 RA Longline M.S., Fares H., Pringle J.R.;
 RT "Role of the yeast Gln4p protein kinase in septin assembly and the
 relationship between septin assembly and septin function.";
 RL J. Cell Biol. 143:719-736(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Dietrich F.S., Mulligan J., Allen E., Atauljo R., Aviles E., Berno A.,
 RA Carpenter J.S., Chen E., Cherry J.M., Chung E., Duncan M.,
 RA Hunnicke-Smith S., Hyman R., Komp C., Laekari D., Lew H., Lin D.,
 RA Mossdale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
 RA Roberts A., Schramm S., Schroeder M., Shogren T., Shroff N.,
 RA Winant A., Yelton M., Botstein D., Davis R.W.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 CC !- FUNCTION: May play a role in septin assembly.
 CC !- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. NIM1
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U33140; AAA75513.1; -
 CC EMBL; U33057; AAB64949.1; -
 CC PIR; S59359; S59359.
 CC HSP; O63450; 1A06.
 CC GeneOnline; 140999; -
 CC SGD; S0002915; GIN4.
 CC GO; GO:0005935; C:bud neck; IDA.
 CC GO; GO:0007117; P:bud growth; IGI.
 CC GO; GO:0001135; P:septin checkpoint; IGI.
 CC GO; GO:0000921; P:septin ring assembly; IGI.
 CC InterPro; IPR000719; Prot_kinase.
 CC InterPro; IPR008271; Ser_thr_kinase.
 CC InterPro; IPR002290; Ser_thr_kinase.
 CC InterPro; IPR001245; Tyr_kinase.
 CC Pfam; PF00069; pkinase; 1.
 CC PRINTS; PD00109; TYRKINASE.
 CC ProDom; PD000001; Prot_kinase; 1.
 CC SMART; SM00220; S_TKG_1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC Transferrase; Serine/threonine-protein kinase; ATP-binding.
 CC DOMAIN 19 289 PROTEIN_KINASE.
 CC FT NP_BIND 25 33 ATP (BY SIMILARITY).
 CC FT BINDING 48 48 ATP (BY SIMILARITY).
 CC FT ACT_SITE 156 156 BY SIMILARITY.
 CC SEQUENCE 1142 AA; 129857 MW; EC16FF4BB49DD811 CRC64;
 SQ
 Query Match 4.3%; Score 176.5; DB 1; Length 1142;
 Best Local Similarity 20.8%; Pred. No. 0.0037;
 Matches 152; Conservative 102; Mismatches 264; Indels 213; Gaps 35;
 QY 41 DRVLEKELQLOP-----TVVALSYINRFTDARREQSLKKKIQLSLTSS-SVS 93
 DB 337 DDEGI-KEKLREPAGANAEKTLVALLY--RPFKCDT---QKELIKQQQVKRQSSVSVS 389
 QY 94 RG-NVSTPP-----RHSSGSFPPVTPPI-----TPSSFRSSTPTEPTQIVRFLRTR 140
 DB 390 PSKKVSTTPQRRNRRESILVTSRKKPIFNKFTASSASSNLTPGSSRRLSKNPSK 449
 QY 141 NLIAPIFLHRLTYMHSNRSTNI--KSKTKFVDDMLSKVPMKGEQS----- 187
 DB 450 KKLSTIVNQSSPTPAS-RNKRAASVINVEKNQKRASIFSTTKNKRSSRSIKRMSLIPTSMK 508
 QY 188 -HLSAHLQLFTGTGFHKNDKRPSPNESEQNSVILEVLVKV-----CHKKRVKQVSC 238

Db 509 RESVTTKLMSTYAKLAEDDDWEYIEKTRKTSNFATLIDEIFEYKYEQIRKEKBLER 568
 QY 239 PIQVPTG-----KQVPL-IPDLNO----- 258
 Db 569 KVREAKARELERRRRKQBEKARKLLEKDKRQBELKQIEIDISOLEBELSKHKE 628
 QY 259 -TXPGNPPSLAVSNEFEPSNSHVMKYSLLFR-----VTRPGRREFNGMNGETNENI 311
 Db 629 EKLGNIRISAPWENEENKINHLVDIDNILRRRNFSLQTRPVSLDPCGMFSSPTTEV 688
 QY 312 DVNEELPARKNRNEDGKTFVQM-----TVFDKNRRLQLLDGEYE--VAMQEM-EE 361
 Db 689 SPIVE--PKRTENERLTTEKKILETIRRSKPLGSGFNIDKELKSKWEYPSIIAPQLSSE 746
 QY 362 CPISKKRAWTETIL--DGKRLPPFE--TPSQOFTLQFTLRWTGTETNDKSTADIAKPLAT 416
 Db 747 RVVSDNSNDGYESILLPKDGNGVSQLKSDTATAPVSDGRLKISE-----IRVPQFT 798
 QY 417 RNESLHOENKPGSVKPTQTIAVKESLTTLQTRKCKDTPNENRQKRLRYQFLYNNTR 476
 Db 799 RKRHFSESNNKRLSV--LSMYSTKESFTNLVDILKGNLDVNNQQSQRI----- 845
 QY 477 QQTAEARDLHCPWCTLNCRLKLYSLLKHLKLCRSRFTFNVYHPKRGARIDVSIKEDVGSY 536
 Db 846 PTPRSADD-----SEFLFE-----TVNE--EAEY 867
 QY 537 AGNPQDIHR-----QPGFA--FSRNGPVKRTPTITHILVCRPKRTKASMSE 579
 Db 868 TGNSSNDRLYDVGDSTIKDSALKLNFAFRNGSNEAKQTDLNHLPLPPLNGDNLK 927
 QY 580 FLESEDEVEQORTYS---SGHNRLYFHSDDTCLPLRQEMEVDSDEKQPEW----- 628
 Db 928 -QNSQEGDQHPKIKSWIPESGSS---HTE-----KEENEKEEKKPEQHKQEDQ 976
 QY 629 -LREKTIITQIE 638
 Db 977 EKREKVDDME 987
 RESULT 10
 Y373 BOVIN
 ID Y373 BOVIN STANDARD; PRT; 1453 AA.
 AC Q9TU23;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein KIAA0373.
 GN KIAA0373.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Trachea;
 RA Jovov B., Ripoll P.J., Benos D.J.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF176816; AAF00990.1; -
 CC Hypothetical protein; Coiled coil.
 KW DOMAIN 37 1426 COILED COIL (POTENTIAL).
 FT DOMAIN 37 1426
 SQ SEQUENCE 1453 AA; 169934 MW; 03CBA0A2A64CF4139 CRC64;

Db 1229 RT-----TLRYDEATYNDKIKE-----KYIAYHKE-----VNQLLKMT 1266

Qy 442 SLTTLQTRKEDTNNRQKLIIFVQFLYNNNT-----RQTEARDLHCP-----WCTL 492

Db 1267 WDTDEYDRKEID-----PKRVNSMFIENKRDGTHKARFVAGDQHPDPTDGMQS 1320

Qy 493 NCRKLYSLKHLKCHSRFIFVYVHPKARDVINECYDGSAGNPDQIHRQGPAPS 552

Db 1321 NVHYHMTSLSLALDN---NYI---TQDIS---SAY---LYADIKEELVIRPPHGL 1369

Qy 553 RNPGRVKTPIHILVCRPKRTKASMEFLES---EDGEVEQORTYSSGHNRLYFHSDFCL 609

Db 1370 MNDKLIRLKKSHVL---KQSGANVYETIKSYLIKQCGMEVGRWSC---VFKNSQVTI 1422

Qy 610 PLRPQEMEVDSDEKDPWLREKTIQIEEFSD---VNEGEKE 649

Db 1423 CLFVDDMILFSKDLN---ANKKIITLKKQYDTRIINLGESD 1461

RESULT 12

SPOT_SCHPO STANDARD; PRT; 1957 AA.

AC Q10411; Q9USE9;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Sporulation-specific protein 15.

GN SPO15 OR SPAC1F3.06C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]_TaxID=4896;

RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.

RX MEDLINE=20107136; PubMed=10639340;

RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;

RT "S. pombe sporulation-specific coiled-coil protein spo15p is localized to the spindle pole body and essential for its modification";

RL J. Cell Sci. 113:545-554(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feitwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares S., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volkart G., Aert R., Robben J., Grynoprez B.,

RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Gcfeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,

RA Galibert F., Aves S., Xiang X., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe";

RL Nature 415:871-880(2002).

RN [3]

RP SEQUENCE OF 705-871 FROM N.A.

RP STRAIN=968 h90;

RX MEDLINE=20223869; PubMed=10759899;

RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,

RA Hiraoka Y.;

RT "Large-scale screening of intracellular protein localization in living fission yeast calls by the use of a GFP-fusion genomic DNA library.";

RL Genes/Cells 5:169-190(2000).

CC -!- FUNCTION: Has a role in the initiation of spore membrane formation.

CC -!- SUBUNIT: Monomer.

CC -!- SUBCELLULAR LOCATION: Spindle pole body.

CC -!- SIMILARITY: Belongs to the MPC70 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; Z70690; CA94624.1; -.

DR EMBL; AB027811; BAA87115.1; -.

DR PIR; T38077; T38077.

KW GeneDB SPombe; SPAC1F3.06C; -.

KW Sporulation; Coiled coil.

FT DOWAIN 199 785 COILED COIL (POTENTIAL).

FT DOWAIN 804 1235 COILED COIL (POTENTIAL).

FT DOWAIN 1320 1471 COILED COIL (POTENTIAL).

FT DOWAIN 1481 1723 COILED COIL (POTENTIAL).

SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA0617D9DA CRC64;

Query Match 3.7%; Score 152; DB 1; Length 1957;

Best Local Similarity 18.5%; Pred. No. 0.22;

Matches 154; Conservative 146; Mismatches 313; Indels 220; Gaps 35;

Qy 25 TLADLIEHIEDHIDTPVLEKQE-----LQPTVVALSYINRPMFTAAR--R 71

Db 833 TLIDNVQKLMHKEVQSKSELKGVNKLSDLKRLNLSLVAISNDQILQLAELSK 892

Qy 72 EQSLKKKIQPKLSLTSSSVSRGNVSTPPRHSG-----SLTPVTPPTPSSSR 123

Db 893 NYDSLEQ-----SAQLNSGLKSLKLEAKQLLHTEENELHRLDKLTKLIEKSKSLDLG 947

Qy 124 SSTPTPTQIYRFTIRNLIAPIFLHRTLTYSMHRNSRTNIKRTKFKVDMLSKVEKMG 183

Db 948 KLTARQEELN-LKEEN-----MSQQAITSVKs--KLDETLSKSKLEA 990

Qy 184 EQES-HLSAHLQITFTGTFPHKNDKPSNSEN-ONSVTLEVLVVKVCHKR---KDVSC 238

Db 991 DIEHLKNKVEVEVERNALLASNERLMDLLKNNGENIASLQTEI---EKKRAENDDLQS 1046

Qy 239 PIQVPTGKQVPLIPD-----LNQTK--PGNFPPLAVSSNFEPSNSHMKVSYSL 287

Db 1047 KLSVSSVEYNLLLSQTKSLEDKTNQLKYEKKVQKLLDEKQNVLEELTSKYK 1106

Qy 288 LFRVTRPGRREFNGMNGETNENIDVNEELPARKENREDGE--KTFVAQMTVFDKNRL 345

Db 1107 L-----GE--ENAIKDELLALRKKSKQHDLCAFPVDDLK--EKSDAL 1146

Qy 346 QLLDGEVAVQMEBCPISKATWETILDGRLPPFTFSOGPTLOFTLRWTGETNDK 405

Db 1147 EQITNEKNELIVSLEQ-----SNSNNE 1168

Qy 406 STAPIAKPLATRNSE---SLHQENKPGSVKPTQIAVKESLTTLDTOTRKEKDTFNENRQK 462

Db 1169 ALVEERSDLANRLSDMKSLSDSDNVISVIRSLRVNDELDT--LKKDKDS----- 1218

Qy 463 LRIFYQLYNNTRQQTTEARDLHCPWCTLNCKRLKSLKHLKCHSRFIFVYVHPK-- 520

Db 1219 -----LSTQYSEVCQDRDD-----LLDSLKGCBSFP-NKYAVSLREL 1254

Qy 521 --GARIDVSNICYDGSY---AGNPDIHR-----QPQF----- 549

Db 1255 CTKSEIDVSEILDDNEFVNFAGNFSLSRLTVLSLENLYLDAFNQVAFKMKMLDNLRTTT 1314
Qy 550 --AFSR-----NGPVKRTPTTHLVCRPK---RTKASMEFFLES-EDGVEQ 590
Db 1315 DAETFKVADLEKLUQEHDDMLIQGDEKALKDSEKFNFLKEAEHTENIHSLEEGKEET 1374
Qy 591 QRTYSSGHRNLYRFSHDTCLPQPMEDVSEDEKPEWLREK--TITQIEFSDVNEGEK 648
Db 1375 KKEIAELSLREDNQLATNKLQDLNQLBIRLKEDVLKESLIISLSE-SLSNQROK 1433
Qy 649 EVMKLNHLVHKGFADNQNHACMLFVENYGGQIKIKNL-CRNFMLHLVSMEDFNLS 707
Db 1434 ESSLLDAXNLEHMLDDTSRKNSSLMKIESINSLODKSFELASAVEKLGALQKLHSES 1493
Qy 708 IMSIDKAVTKLREMOQKLEKESGSA-SPANEBITEQNGTANGFGEINSKSKAL 759
Db 1494 LSLMENIKSQLEAKEIKQVDESTIQELDHEITASKN---NYECKLNDKDSII 1543

RESULT 13

UBP3 YEAST STANDARD; PRT; 912 AA.
AC Q01477;
DT 01-JUL-1993 (Rel. 26, Created)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 3 (EC 3.1.2.15) (Ubiquitin
thiolesterase 3) (Ubiquitin-specific processing protease 3)
DE (Deubiquitinating enzyme 3).
GN UB3 OR YBR151C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN [2]
RN [3]
RX MEDLINE=93054674; PubMed=1429680;
RA Baker R.T., Tobias J.W., Varshavsky A.;
RT "Ubiquitin-specific proteases of Saccharomycetes cerevisiae. Cloning of
UBP2 and UB3, and functional analysis of the UB3 gene family.";
RL J. Biol. Chem. 267:23364-23375 (1992).
RC STRAIN=S288C / AB972;
RX MEDLINE=9313264; PubMed=9169868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomycetes cerevisiae chromosome V.";
RL Nature 387:78-81 (1997).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=96361363; PubMed=8752220;
RA Moazed D., Johnson D.;
RT "A deubiquitinating enzyme interacts with SIR4 and regulates
silencing in S. cerevisiae.";
RL Cell 86:667-677 (1996).
CC -!- FUNCTION: Has an ATP-independent isopeptidase activity, cleaving
at the carboxyl terminus of the ubiquitin moiety in natural or
engineered linear fusion proteins, irrespective of their size or
the presence of an amino-terminal extension to ubiquitin.
CC -!- FUNCTION: Plays a role in regulation of silencing by interacting
with SIR4.
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC This SWISS-PROT entry is covaried. It is produced through a collaboration

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M94917; AAA35191.1; -;
CC EMBL; U18917; AAB64678.1; -;
CC PIR; B44450; B44450.
CC GERMOnline; 139229; -;
CC MEROPS; C19.004; -;
CC SGD; S0000953; UBP3.
CC GO; GO:0005737; Cytoplasm; IC.
CC GO; GO:0004843; Fubiquitin-specific protease activity; IDA.
CC InterPro; IPR001394; Peptidase_C19.
CC Pfam; PF00443; UCH_1.
CC PROSITE; PS00972; UCH_2_1; 1.
CC PROSITE; PS00973; UCH_2_2; 1.
CC PROSITE; PS02355; UCH_2_3; 1.
KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT ACT SITE 469 469 BY SIMILARITY.
FT ACT SITE 852 852 BY SIMILARITY.
FT ACT SITE 861 861 BY SIMILARITY.
SQ SEQUENCE 912 AA; 101916 MW; 337FD05527C5542D CRC64;

Query Match 3.7%; Score 150; DB 1; Length 912;
Best Local Similarity 18.5%; Pred. No. 0.11;
Matches 153; Conservative 126; Mismatches 287; Indels 260; Gaps 37;
Qy 65 MTDAAARREQESLKKIQLPKLSLTSSVSRRGNVSTPPHSSGSLPPVTPPTSSSPRS 124
Db 3 MQDANKESYSNYPK-----TSSPPPTPTNMQIPYQAPLQWGYTQ 45
Qy 125 STPTPTPTQI--YRFLRTNLI---APIFLHRT-----LTYMSHN-----SRTNIKRTP 169
Db 46 APYLYPTQIPAYSF---NMVNQNFY-HQSGSPHLPQNNINGSGSTNNNNKKNK 100
Qy 170 KYDDMLSKVKMKKGQESHLSAHLQLTFTGFHNDK-----PSNS-----E 213
Db 101 HNGITNN-NGSGNGQAGNSGSGMYNKSHTYHHYNNHNPMMASNSNSAGMKQT 159
Qy 214 NEQN-----SVTLEVLVVKVCHKKRDXPCIRQVPTGKKQVPLPDLNQT 259
Db 160 NSSNGNGSSATSPSYSSNSQYDLKFDVTKLN-----LKNSSNLIQLPLF--INTT 213
Qy 260 KPG-----NFPSLAVSSNEFEPN-----SHMVKSYLLFRVTRPGREFNGM 302
Db 214 EAEFAAASVQVELAKNLNLSLENSLENSVKSAAHHTKSHSI-----PQNEE 264
Qy 303 INGETN-ENIDVNEELPARRKENRDEGKTFVAQMTVPDKNRRLQLLDGEYEVANQEMEE 361
Db 265 VKTETHGEEDAHDKPHASKDAHELKKTEVKK-----EDAKQDRNE 307
Qy 362 CRISKGRATWEILDGKRLPPPTFSQGTFTLQWGTGETNDKSTAPIA-----KPLATR 417
Db 308 KVIQEPQATVLPVVDKKE--PEESVEEN-----TSKTSPPSPSPAASWASAIASD 356
Qy 418 NGSLSHOENKPGSVKPTQITIAVKESLITDLOTRKEKPTNENRQKIRIFYOFLYNNNRQ 477
Db 357 AIKSRQASNKTVSGSVWTKPTISGTTAGVSSINMAATIGKSSPL-----LSK 405
Qy 478 QTEARDLHCPMCTNCRKLYSLKHLKCHGRFTFNVTYHPKGARIDVSNCEVDGYA 537
Db 406 QPKKKDKKTVPPSTKGIEPLGSIA--LRMCFDPDFISYVLNRKNKVENKIPVHSII----- 458
Qy 538 GNPQDHRPGGFAFRNGPVKPTPIHILVCRP-----KTKASM----- 577
Db 459 --PRGLINRANICF-----MSSVLQVLLYCKPFDIVNLVSTRNTNSRVGTSSCKLLDA 510
Qy 578 -----SEFLESDG--EVEQRTYSSGHRNLYPHSDTCLPLRPQNEVDSE 621
Db 511 CLTMVKQFDKTYKKFLENLADDAKTTESDAKSKSKSFOHCATADAVPDEF----- 565

| | | | |
|----|-----|--|-----|
| QY | 622 | DEKDPWLREXTIQIEEFSVNGEVEVMKLNHLVWCH-----GFTADNQMHQA | 572 |
| Db | 566 | -----YKLSITPKFDQLQWGHGDEAEFTLLDQLHBEILSALDGTGDNE---- | 612 |
| QY | 673 | CMLFVENVGQKIIKNLCRNFMHLVSMH---DFNLISIMSIDKATVKLREWQOKLEKGE | 729 |
| Db | 613 | -----IQNMLQINDEQL-KYFFIRNLGRYKAEF-----IKNASPRKELIEKY---- | 656 |
| QY | 730 | SASPAÑEEITQONGTANGSEINSKEKALETDSVGVSQSKKQK | 775 |
| Db | 657 | --GVINDDSTEE-----NGWHE-----VSGSKRGKHKTK | 693 |

RESULT 14

| | | | |
|----|--|-----------------------------------|---------------|
| ID | USOL_YEAST | STANDARD; | PRT; 1790 AA. |
| AC | P25386; | | |
| DT | 01-MAY-1992 | (Rel. 22, Created) | |
| DT | 01-MAY-1992 | (Rel. 22, Last sequence update) | |
| DT | 15-MAR-2004 | (Rel. 43, Last annotation update) | |
| DE | Intracellular protein transport protein USOL1. | | |
| GN | USOL1 OR INT1 OR YDL058W. | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomycetes. | | |
| OX | NCBI_TaxID=4932; | | |
| | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=X2180-1A; | | |
| RC | MEDLINE=91185402; PubMed=2010462; | | |
| RA | Nakajima H., Hirata A., Osawa Y., Yonehara T., Yoda K., | | |
| RA | Yamasaki M.; | | |
| RT | "A cytoskeleton-related gene, usol1, is required for intracellular | | |
| RT | protein transport in Saccharomyces cerevisiae."; | | |
| RL | J. Cell Biol. 113:245-260(1991). | | |
| RN | [2] | | |
| RP | SEQUENCE OF 782-1790 FROM N.A. | | |
| RA | Hoschter M.K., Herman D.J., Bendel C.M., McLeellan M., Tao N., | | |
| RA | Kendrick X.E.; | | |
| RL | Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases. | | |

```

RP SEQUENCE OF 1-8 FROM N.A.
RL Bai I., Symington L.S.;
RA Submitted (MAY-1996) to the EMBL/genBank/DBJ databases.
CC
CC -!- FUNCTION: Required for protein transport from the ER to the Golgi
CC complex.
CC
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, composed
CC of an heptapeptide repeat pattern characteristic of alpha-helical
CC coiled coils. May form filamentous structures in the cell.
CC
CC -!- SIMILARITY: BELONGS TO THE VDP/USO1/YSL047C FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC

```

```

CC
-----
EMBL; X54378; CAAB38253.1; -
EMBL; L03188; AAB00143.1; -
EMBL; U53668; AAB66659.1; -
GeneOnline; I40300; -
SGD; S0002216; US01.
InterPro; IPR008938; ARM.
InterPro; IPR02017; Spectrin.
InterPro; IPR008955; Usl1_p115_C.
InterPro; IPR008953; Usl1_p115_head.
Pfam; PF04871; Usl1_p115_C; 1.
Pfam; PF04869; Usl1_p115_head; 1.
DR
DR
DR

```

QY 725 LEK-----GESASPAHEITEEQNGTANGFSEINSKEKALE- 760
 DB 1510 LESTIESNETELKSMETIRKSDKLEQSKASABEDIKNLQHEKSDLSIRINSEKDIIE 1569
 QY 761 -----TDSVSGVSKQSKQL 776
 DB 1570 LKSKLRIEAKSGSELETVKQEL 1591

RESULT 15
 ID GOA4 HUMAN STANDARD; PRT: 2230 AA.
 AC Q13439; Q13270; Q13654; Q14436;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2003 (Rel. 42, Last annotation update)
 DE Golgi autoantigen, golgin subfamily A member 4 (Trans-Golgi p230) (256
 DE xDa golgin) (Golgin-245) (72.1 protein).
 GN GOLGA4.
 OS Homo sapiens (Human)
 OC Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RA MEDLINE=96215236; PubMed=8626529;
 RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
 RT "Molecular characterization of trans-Golgi p230: a human peripheral
 RT membrane protein encoded by a gene on chromosome 6p12-22 contains
 RT extensive coiled-coil alpha-helical domains and a granin motif.";
 RL J. Biol. Chem. 271:8328-8337(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Seelig H.P.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 131-2230 FROM N.A.
 RC TISSUE=Placenta;
 RA MEDLINE=96125112; PubMed=8537393;
 RA Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan B.K.L.;
 RT "Molecular characterization of golgin-245, a novel Golgi complex
 RT protein containing a granin signature.";
 RL J. Biol. Chem. 270:31262-31268(1995).
 RN [4]
 RP SEQUENCE OF 524-672 FROM N.A.
 RC TISSUE=Gastric fundus;
 RA Balague C.;
 RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
 CC -!- FUNCTION: May play a role in vesicular transport from the trans-
 CC Golgi.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein
 CC associated with the trans-Golgi network.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q13439-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q13439-2; Sequence=VSP_004272, VSP_004273;
 CC Name=3;
 CC IsoId=Q13439-3; Sequence=VSP_004274;
 CC Name=4;
 CC IsoId=Q13439-4; Sequence=VSP_004275;
 CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
 CC -!- DISEASE: Antibodies against GOLGA4 are present in sera from
 CC patients with Sjogren's syndrome (SS) [MIM:270150]. Sera from
 CC patients with SS often contain antibodies that react with normal
 CC components of the Golgi complex.
 CC -!- DISEASE: Antibodies against GOLGA4 are found in sera from
 CC hepatitis B patients.
 CC -!- SIMILARITY: Belongs to the golgin family.
 CC -!- SIMILARITY: Contains 1 GRIP domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U41740; AAC50434.1; -;
 DR EMBL; X82834; CAA58041.1; -;
 DR EMBL; U31906; AAC51791.1; -;
 DR EMBL; X76942; CAA54261.1; -;
 DR Genew; HGNC:4427; GOLGA4.
 DR MIM; 602509; -;
 DR MIM; 270150; -;
 DR GO; GO:0003802; C:Golgi trans face; TAS.
 DR GO; GO:0016192; P:vesicle-mediated transport; TAS.
 DR InterPro; IPR000237; GRIP_domain.
 DR Pfam; PF01465; GRIP; 1.
 DR PROSITE; PS00913; GRIP; 1.
 KW Antigen; Golgi stack; Coiled coil; Alternative splicing.
 FT DOMAIN 133 2185
 FT DOMAIN 252 2096
 FT DOMAIN 2168 2215
 FT VARSPLIC 2154 2185
 FT AICTRMGSHLENLPNLSICEKCLSI (in isoform
 FT 2).
 FT /FTId=VSP_004272.
 FT Missing (in isoform 2).
 FT /FTId=VSP_004273.
 FT Missing (in isoform 3).
 FT /FTId=VSP_004274.
 FT FTSPRSRGIF--> SWLRSSS (in isoform 4).
 FT /FTId=VSP_004275.
 FT R-> K (IN REF. 3).
 FT Y-> H (IN REF. 3).
 FT T-> A (IN REF. 3).
 FT K-> E (IN REF. 3).
 FT T-> A (IN REF. 3).
 FT K-> E (IN REF. 3).
 FT K-> N (IN REF. 3).
 FT CONFLICT 682 682
 SQ SEQUENCE 2230 AA; 261139 MW; 3BB733DBLEA86134 CRC64;
 Query Match 3.6%; Score 147; DB 1; Length 2230;
 Best Local Similarity 18.2%; Pred. No. 0.53;
 Matches 152; Conservative 148; Mismatches 308; Indels 228; Gaps 36;
 QY 30 IEHIEDNHDTPRVLEK-----QELQOPTYVALSVINREMTDAARQESLKKIQ 81
 DB 330 LEKIKDLHMAETKLIQLRDAKNLIEQEQKGVIAETKQMHETLEMEEEIAQLRS 389
 QY 82 PKLSLTSSSVSRGNVSTPPRHSSGLTPPVTPPTPSSPSFSSPTPTPTIYRFLRPN 141
 DB 390 RIQMTTQGEELREQEKESRAAFBELEKALSTAQKTEARAKLKAEMDEQIKTEKTS 449
 QY 142 LIAPFLHRTLTYMSHRNSRTNIKRTKV-----DDMLSKVEKM-----KGQESHSL 190
 DB 450 -----EERISLQQLSRVQGVVDVVMKSSQEQIALQKLHEKELARKQGE---L 496
 QY 191 SAHLQLTFTGFFHKNDKPSNSENQNSVTLEVL---LVKVCCHKRKRKDVCFIRQVPTGK 247
 DB 497 TKKLTQTRREF-----QEQMKVALEKSQSEYVKLSQBEKQEQESLALBELELQK 544
 QY 248 KOV-----PLPDNLQTPGNPPPSLAVSSNPFPSNSHMVKSYSLLFVTPGPRREFNGM 302
 DB 545 KAILTESENKURDLQ-----EAETRYTRILESSLESLSKQENKQSKDL 590
 QY 303 ---INGETNENIDVNEELPARRKRNREDGEKTFVAQMTVFDKNRRLQLLDGEYEYVQMGM 359
 DB 591 AVHEAEKMKH---NKEITVAVVEKHKTELESIAKHQODALW---TEKLVQLKQYQTEMEKL 645
 QY 360 -BECPISKRATWETILDGKLLPPTFSQPTQLQTLRTWTGETNDSKTAPI-----AK 412

Db 646 REKCEQEK-----ETLLKDKBI-IFQAHIE-----EMNEKTLEKLDVKQTELE 687
QY 413 PLATRNSSL---HGENKPGSVKPTQTIAVKESLTTDLQTRKEKDTNENQKLRIFYQF 469
Db 688 SLSSELSVLKARHLEELSVLKQDTDKMKQELEAKGDEOK-----NHQOQVDSIIKE 742
QY 470 LYNNTRQOTEARDLHCPWCTLNCRLKLSLL-----KHLKLSRFPFIPVYHPKGARID 525
Db 743 HEVSIQRT'EKALKDQIN-----QLELLKRDKHLK-----EQAHVENLEAD 785
QY 526 VSINECYDGSYAGNPQDIHQPGFAPSRNGPVPKTPITHILVCRPKRTKASMSF----- 580
Db 786 IKRSE-GELOQASAKLDVFSYQSA-----TH-----EQTKAYEOLAOLOQ 826
QY 581 ----LESED-----GEVEQORTYSSGHNRLYFHSDTCLPLRPQEMEVD-----EDEK 624
Db 827 KLLDLETERILLTKQVAEVAQK-----KDVCTELDAHKIQVQDLMOQLEKON 874
QY 625 DPWLRKKTITQIEFSDVNEGEKEVMKWLHVMKHGFIADNOMNHACMLFVENYQKI 684
Db 875 SEMEQKVKSLTQVYE-SKLEDGNKEQEQTKOILVEKENMIL--QVREGQKKEIILTQKL 931
QY 685 IKKNLCRNFMHLHVSMDPNLISIMSIDKAVTKLREMOQKLEKGESASPANEE----- 737
Db 932 SAKE-----DSIHILN-----ZEYETKPKNQEKMEKVKQAKENQETILKKLIL 975
QY 738 -----ITEEQNGTANGFSEINSKEKALE-----TDSVSGVSKSQKQ 774
Db 976 DOEAKLKKLENTA---LELSQKEKQFNAKMLEMAQANSAGISDAVSRLETNQKEQ 1028

Search completed: August 25, 2004, 17:55:08
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:58:25 ; Search time 132 Seconds
(without alignments)
1661.037 Million cell up

Title: US-09-874-162A-8

CS-03-074-102A-0
 Date: 776
 Perfect score:
 Sequence: 1 MTGIAAASFNTCRFGCG.....KALETDSVSGVSKOSKKKL 776

Scoring table: OLIGO
Gapop 60.0 Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

```
Database : A_Geneseq_29Jan04.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003a.*
7: geneseqp2003b.*
8: geneseqp2004.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | | | DB | ID | Description |
|------------|-------|-------|--------|----|----------|--------------------|
| | Score | Match | Length | | | |
| 1 | 647 | 83.4 | 803 | 3 | AAB00066 | Aa000066 KIAA0160 |
| 2 | 355 | 45.7 | 739 | 5 | ABG30842 | Abg30842 Human Jci |
| 3 | 346 | 44.16 | 776 | 5 | ABG30843 | Abg30843 Human Jaz |
| 4 | 283 | 36.5 | 289 | 4 | AAU15958 | Aau15958 Human nov |
| 5 | 283 | 36.5 | 289 | 6 | ABU55027 | Abu55027 Human nov |
| 6 | 186 | 24.0 | 388 | 4 | AAU15978 | Aau15978 Human nov |
| 7 | 186 | 24.0 | 388 | 6 | ABU55047 | Abu55047 Human nov |
| 8 | 161 | 20.7 | 292 | 4 | AAU16416 | Aau16416 Human nov |
| 9 | 161 | 20.7 | 292 | 6 | ABU55485 | Abu55485 Human nov |
| 10 | 131 | 16.9 | 175 | 4 | AAU16402 | Aau16402 Human nov |
| 11 | 131 | 16.9 | 175 | 6 | ABU55471 | Abu55471 Human nov |
| 12 | 128 | 16.5 | 243 | 7 | ADBe4250 | Adbe4250 Human pro |
| 13 | 128 | 16.5 | 243 | 7 | ADCO6802 | Adco6802 Human pro |
| 14 | 109 | 14.0 | 278 | 4 | AAU16381 | Aau16381 Human nov |
| 15 | 109 | 14.0 | 278 | 6 | ABU55450 | Abu55450 Human nov |
| 16 | 93 | 12.0 | 143 | 4 | AAU15931 | Aau15931 Human nov |
| 17 | 93 | 12.0 | 143 | 6 | ABU55000 | Abu55000 Human nov |
| 18 | 91 | 11.7 | 243 | 5 | ABG30841 | Abg30841 Human Jux |
| 19 | 86 | 11.1 | 202 | 3 | AAB43387 | Aab43387 Human ORF |
| 20 | 56 | 7.2 | 171 | 4 | ABBI1690 | Abbi1690 Human sfp |
| 21 | 37 | 4.8 | 102 | 3 | ABG03182 | Abg03182 Human sec |
| 22 | 14 | 1.8 | 14 | 4 | AAK97551 | Aak97551 Human pep |
| 23 | 11 | 1.4 | 955 | 4 | ABB71635 | Abb71635 Drosophil |
| 24 | 9 | 1.2 | 96 | 6 | ADA33026 | Ada33026 Acinetoba |
| 25 | 9 | 1.2 | 1991 | 4 | ABBS0651 | Abbs0651 Drosophil |

ALIGNMENTS

RESULT 1

RESUL I
AAB00066
ID AAB00066 standard: protein: 803 AA:

XX
AC AAB00066:

16-NOV-2000 (first entry)

XX DE KTAA0160 polimentide

XX Vernalization gene; VRN2; plant characteristic; flowering time;
KW leaf size; leaf shape; shade avoidance response; reproduction; breeding;
KW pollination; cultivation; human.

XX
OS
Homo sapiens.

XX PN WO200044918-A1

03-AUG-2000
PD
XX

XX.
28-TAN-2000: 2000WO-GB000248

XX
BP 39-JAN-1999. 99CB-00001937

XX
 (PLANT BIOSCIENCE LTD)

XX
PT
Dean C
Gendall A.

XX
WPT. 2000 100233/11

DR N-PSDB; AAA47759;

Isolated vernalization gene VRN2 is used to produce transgenic plants with altered vernalization response, flowering time, leaf size and/or shape or shade avoidance response for maximized reproductive success.

PS Disclosure: Page 79: 105pp: English.

Isolated nucleic acid sequences obtained from the VRN2 locus of a plant encode polypeptides which are capable of affecting one or more vernalization responses such as, flowering time, leaf size and/or shape or the shade avoidance response of a plant into which the nucleic acid is introduced. Introducing such sequences into plants to alter these characteristics maximises the reproductive success of the plant. This protein is encoded an isolated human sequence which has homology to the Arabidopsis thaliana VRN2 gene over a short region near the N-terminus

Sequence 803 AA;

Sequence 803 AA;

RESULT 3
 ABG30843
 ID ABG30843 standard; protein; 776 AA.
 XX
 AC ABG30843;
 XX
 DT 21-OCT-2002 (first entry)
 DE Human JAZF1/JJAZ1 fusion protein.
 DE
 XX
 KW Human; JAZF1; juxtaposed with another zinc finger; JJAZ1; JAZF1/JJAZ1;
 KW joined with JAZF1; proliferation; endometrial stroma tumour; immunogen;
 KW antigen; antibody; fertility; pregnancy; gene therapy; vaccine;
 KW fusion protein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS Chimeric.
 XX
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 13..37
 FT /note= "Encoded by ACCTGCCGATCGGGGGCTCGGACTCCATTC-
 FT CCCACCTCGCGGACCTCATCGACATCGAGGACCAACAC. C2H2 zinc
 FT finger domain"
 FT 99..100
 FT /note= "Breakpoint for production of fusion protein
 FT JAZF1/JJAZ1"
 FT Misc-difference 357
 FT /note= "Encoded by CAG"
 FT Misc-difference 485..508
 FT /note= "Encoded by CTGCATTGCTTGGTGTACTCTGAACTGCCGAA-
 FT ACTTATAGTTTACTCAAGCATCTTAACCTGCGCAT. C2H2 zinc finger
 FT domain"
 FT 558..575
 FT /note= "Bipartite nuclear localisation signal"
 XX
 PN WO200193805-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 04-JUN-2001; 2001WO-US017936.
 XX
 XX
 PR 02-JUN-2000; 2000US-0209093P.
 XX
 XX (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
 XX
 XX Koontz J, Sklar J;
 XX
 DR WPI; 2002-575047/61.
 DR N-PSDB; ABR89163.
 XX
 XX Novel JAZF1, JJAZ1 or JAZF1/JJAZ1 polypeptides useful as immunogens or
 PT antigens to raise or test anti-JAZF1, JJAZ1 or JAZF1/JJAZ1 antibodies.
 XX
 XX Claim 7; Fig 3; 76pp; English.
 XX
 CC The present invention relates to a new JAZF1 (juxtaposed with another
 CC zinc finger), JJAZ1 (joined with JAZF1) or JAZF1/JJAZ1 polypeptide. The
 CC methods of the invention can be used to identify a compound which
 CC controls proliferation of endometrial stroma, by expressing JJAZ1 in the
 CC presence of the compound, and determining whether the compound affects
 CC expression of JJAZ1, JJAZ1 or JAZF1/JJAZ1 polypeptides are useful
 CC as immunogens or antigens to raise or test anti-JAZF1, JJAZ1 or
 CC JAZF1/JJAZ1 antibodies. The invention can be used as bait proteins in a
 CC two hybrid assay or three hybrid assay to identify other proteins which
 CC bind or interact with JAZF1/JJAZ1-binding proteins. JAZF1, JJAZ1 or
 CC JAZF1/JJAZ1 molecules are useful for identifying the origin of tumour and
 CC as tumour marker protein to verify that a stromal tumour is from
 CC endometrium. The antibody is useful for promoting or decreasing fertility
 CC or pregnancy, and also for treating endometrial stromal tumours. The
 CC present amino acid sequence represents the human JAZF1/JJAZ1 fusion
 CC protein of the invention

SQ Sequence 776 AA;
 Query Match 44.6%; Score 346; DB 5; Length 776;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 38 IDTDPVLEKQELQOQPTVALSYINRPMWTDAAAREQESLKKIOPKLSLTSSVSRGNV 97
 DB 38 IDTDPVLEKQELQOQPTVALSYINRPMWTDAAAREQESLKKIOPKLSLTSSVSRGNV 97
 QY 98 STPPRHSSGSLTPVTPPTIPSSFRSSTPTPTQYRFLRNLNLIAPILHRTLTVMWSH 157
 DB 98 STPPRHSSGSLTPVTPPTIPSSFRSSTPTPTQYRFLRNLNLIAPILHRTLTVMWSH 157
 QY 158 RNSRTNKRKTKFKYDDMLSKVKMGQESHSLSAHLQLTFTGFFHKNDKPSNSENEQN 217
 DB 158 RNSRTNKRKTKFKYDDMLSKVKMGQESHSLSAHLQLTFTGFFHKNDKPSNSENEQN 217
 QY 218 SVTLLEVLVVKVCHKKRKDVSCPIRQVPTGKKQVFLPDLNCTYKGNFPPSLAVSSNEPEPS 277
 DB 218 SVTLLEVLVVKVCHKKRKDVSCPIRQVPTGKKQVFLPDLNCTYKGNFPPSLAVSSNEPEPS 277
 QY 278 NSHMVKSYSLFRVTRPGREFNMGINGETNENIDVNEELPARRKRNREDGKTFVAQMT 337
 DB 278 NSHMVKSYSLFRVTRPGREFNMGINGETNENIDVNEELPARRKRNREDGKTFVAQMT 337
 QY 338 VFDKRRQLQLDGEYEVAMQEMEECPISKRRATWETILDGKRLPPPTPSQGTLOFTLR 397
 DB 338 VFDKRRQLQLDGEYEVAMQEMEECPISKRRATWETILDGKRLPPPTPSQGTLOFTLR 397
 QY 398 WTGETNDKSTAPAKPLATRNSESLHQNKPQSVKPTQTIKVESLTTDLQTRKEKDTN 457
 DB 398 WTGETNDKSTAPAKPLATRNSESLHQNKPQSVKPTQTIKVESLTTDLQTRKEKDTN 457
 QY 458 ENQKRLIFYQFLYNNNTROQTEARDD 484
 DB 458 ENQKRLIFYQFLYNNNTROQTEARDD 484
 RESULT 4
 AAU15958
 ID AAU15958 standard; protein; 289 AA.
 XX
 AC AAU15958;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human novel secreted protein, Seq ID 911.
 DE
 XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
 KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
 KW antibacterial; virucide; fungicide; ophthalmological; vulnary;
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;
 KW preservative; antiproliferative.
 XX
 OS Homo sapiens.
 XX
 PN WO200155322-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US001341.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225288P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225577P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.

PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.

(HUNA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-488783/53.

N-PSDB; AAS25945.

New nucleic acid molecules encoding 461 human secreted proteins for diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives.

Claim 11; SEQ ID NO 911; 980pp; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemoaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 36.5%; Score 283; DB 4; Length 289;
Best Local Similarity 100.0%; Pred. No. 6.9e-277; Indels 0; Gaps 0;
Matches 283; Conservative 0; Mismatches 0;
QY 494 CRKLSLLKHLKLSRIFVYVHPKGRIDVSNICVDSYAGNPDIIHQPCGAFPSR 553
DB 7 CRKLSLLKHLKLSRIFVYVHPKGRIDVSNICVDSYAGNPDIIHQPCGAFPSR 66
QY 554 NGPVKRTITHLVCRPKRTKASMEFLESDGEVEQORTYSSGHNRLYFHSDDCLPLRP 613
DB 67 NGPVKRTITHLVCRPKRTKASMEFLESDGEVEQORTYSSGHNRLYFHSDDCLPLRP 126
QY 614 QMEVDSDEKDPWLRKTTITQIEFSDVNEGEKVMKLNHLVHKHGFADNQMNHAC 673
DB 127 QMEVDSDEKDPWLRKTTITQIEFSDVNEGEKVMKLNHLVHKHGFADNQMNHAC 186
QY 674 MLFVNYGQKIIKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGEASGP 733
DB 187 MLFVNYGQKIIKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGEASGP 246
QY 734 ANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776
DB 247 ANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 289

RESULT 5
ABU55027
ID ABU55027 standard; protein; 289 AA.
XX AC ABU55027;
XX DT 18-MAR-2003 (first entry)
XX DE Human novel polypeptide #114.
XX KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX OS Homo sapiens.
XX PN US2002132753-A1.
XX PD 19-SEP-2002.
XX PF 17-JAN-2001; 2001US-00764864.

XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180828P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225367P.
PR 14-AUG-2000; 2000US-0225368P.
PR 14-AUG-2000; 2000US-0225370P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 22-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 03-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.

(ROSE/) ROSEN C A.
(RUBE/) RUBEN S M.
(BARA/) BARASH S C.
Rosen CA, Ruben SM, Barash SC;
WPI; 2003-147444/14.
DR N-PSDB; ABX73286.
New polypeptides and nucleic acids, useful in gene therapy for treating,
inhibiting or preventing e.g. neural, immune system, muscular,
respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
renal disorders.
Claim 11; SEQ ID NO 911; 402pp; English.
The invention relates to human novel polypeptides and their associated
polynucleotides. The polypeptides and polynucleotides are useful in gene
therapy for treating, inhibiting or preventing neural disorders, immune
system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
and multiple sclerosis), muscular disorders, respiratory diseases (e.g.

CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
XX ABU55748 represent human novel polypeptides of the invention
SQ Sequence 289 AA;

Query Match 36.5%; Score 283; DB 6; Length 289;
Best Local Similarity 100.0%; Pred. No. 6.9e-277;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 494 CRKLYSLKHLKCHSRIFNYVYHPKGRIDVINECYDGSYAGNPQDIHQPGAFSR 553
DB 7 CRKLYSLKHLKCHSRIFNYVYHPKGRIDVINECYDGSYAGNPQDIHQPGAFSR 66
QY 554 NGPVKRTPIHLVCRPKRTKASSEHLESEDEGEVQRTYSSGHNLPHSDTCLPLRP 613
DB 67 NGPVKRTPIHLVCRPKRTKASSEHLESEDEGEVQRTYSSGHNLPHSDTCLPLRP 126
QY 614 QMEVDSDEKDPWLREKTTQIEEESDVNVEGEKVMKLNHLVHMKHGFADNQMHAC 673
DB 127 QMEVDSDEKDPWLREKTTQIEEESDVNVEGEKVMKLNHLVHMKHGFADNQMHAC 186
QY 674 MLFVNYQCKIIKKNLCSFNLHLSVMDHFNLSIMSDKAVTKLREMQCKLEKGESASP 733
DB 187 MLFVNYQCKIIKKNLCSFNLHLSVMDHFNLSIMSDKAVTKLREMQCKLEKGESASP 246
QY 734 ANEETEEQNTANGFSEINSKEKALETDSVSGVSKOSKKQKL 776
DB 247 ANEETEEQNTANGFSEINSKEKALETDSVSGVSKOSKKQKL 289

RESULT 6
AAU15978
ID AAU15978 standard; protein; 388 AA.

AC AAU15978;

DT 07-NOV-2001 (first entry)

DE Human novel secreted protein, Seq ID 931.

XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.

OS Homo sapiens.

PN WO200155322-A2.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US001341.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.
PR 27-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244826P.
 PR 08-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249219P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 05-JAN-2001; 2001US-0259678P.
 (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-488783/53.
 XX N-PSDB; AAS25965.
 XX
 XX New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX
 XX Claim 11; SEQ ID NO 931; 980pp; English.
 PS
 XX The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,

CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodysplasia,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Query Match 24.0%; Score 186; DB 4; Length 388;
 Best Local Similarity 99.5%; Pred. No. 1e-178;
 Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 389 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKPGSVKPTQTIAVKESLTDLQ 448
 Db 1 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLSHQENKPGSVKPTQTIAVKESLTDLQ 60
 Qy 449 TRKEKTPNENROKLAIFYOFLYNNTRQQTTEARDLHCPWCTLNCRKLYSLKHLKCH 508
 Db 61 TRKEKTPNENROKLAIFYOFLYNNTRQQTTEARDLHCPWCTLNCRKLYSLKHLKCH 120
 Qy 509 SRPIFNVYHPKGARIDVSNIECYGAGNPDIIHQPGFAFRNGPVKRTPIITLVC 568
 Db 121 SRPIFNVYHPKGARIDVSNIECYGAGNPDIIHQPGFAFRNGPVKRTPIITLVC 180
 Qy 569 RPKRTKASMEFLESEDEGEVEQRTYSSGHNRLYPHRDTCLPLRQMEVDEKDPW 628
 Db 181 RPKRTKASMEFLESEDEGEVEQRTYSSGHNRLYPHRDTCLPLRQMEVDEKDPW 240
 Qy 629 LREKTIQTEERSDVNEGEKEVMKLNLVHVMKHGFTADNMHACMLFVENYQOKIKKN 688
 Db 241 LREKTIQTEERSDVNEGEKEVMKLNLVHVMKHGFTADNMHACMLFVENYQOKIKKN 300
 Qy 689 LCRNFMHLVSMHDFNLISIMSIDKAVTKLRENQOKLEKESASPANEETEPQNGTANG 748
 Db 301 LCRNFMHLVSMHDFNLISIMSIDKAVTKLRENQOKLEKESASPANEETEPQNGTANG 360
 Qy 749 FSEINSKEKALETDSVSGVSKQSKOKL 776
 Db 361 FSEINSKEKALETDSVSGVSKQSKOKL 388

RESULT 7
 ABUS5047
 ID ABUS5047 standard; protein; 388 AA.
 XX
 AC ABUS5047;
 XX
 XX 18-MAR-2003 (first entry)
 DT
 XX
 XX Human novel polypeptide #134.
 DE
 XX Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX

| | | | |
|----|---|--|--|
| OS | Homo sapiens. | PS | Claim 11; SEQ ID NO 931; 402pp; English. |
| XX | | XX | The invention relates to human novel polypeptides and their associated |
| PN | US2002132753-A1. | XX | polynucleotides. The polypeptides and polynucleotides are useful in gene |
| XX | | CC | therapy for treating, inhibiting or preventing neural disorders, immune |
| PD | 19-SEP-2002. | CC | system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis |
| XX | | CC | and multiple sclerosis), muscular disorders, respiratory diseases (e.g. |
| PF | 17-JAN-2001; 2001US-00764864. | CC | nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders, |
| XX | | CC | gastrointestinal disorders, pulmonary disorders, cardiovascular disorders, |
| PR | 31-JAN-2000; 2000US-0179065P. | CC | (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left |
| PR | 04-FEB-2000; 2000US-0180628P. | CC | heart syndrome), renal disorders (e.g. acute kidney failure and end-stage |
| PR | 28-JUN-2000; 2000US-0214886P. | CC | renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and |
| PR | 07-JUL-2000; 2000US-0216647P. | CC | leukemia), inflammatory diseases (e.g. septic shock, bursitis and |
| PR | 07-JUL-2000; 2000US-0216880P. | CC | appendicitis), allergic reactions and conditions (e.g. asthma), blood |
| PR | 11-JUL-2000; 2000US-0217487P. | CC | related disorders (e.g. thrombosis, atherosclerosis and myocardial |
| PR | 11-JUL-2000; 2000US-0217496P. | CC | infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and |
| PR | 14-JUL-2000; 2000US-0218290P. | CC | ABUS5748 represent human novel polypeptides of the invention |
| PR | 26-JUL-2000; 2000US-0220963P. | XX | |
| PR | 26-JUL-2000; 2000US-0220964P. | XX | Sequence 388 AA; |
| PR | 14-AUG-2000; 2000US-0224518P. | SQ | |
| PR | 14-AUG-2000; 2000US-0224519P. | | Query Match 24.0%; Score 186; DB 6; Length 388; |
| PR | 14-AUG-2000; 2000US-0225267P. | | Best Local Similarity 99.5%; Pred. No. 1e-178; |
| PR | 14-AUG-2000; 2000US-0225268P. | | Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0; |
| PR | 14-AUG-2000; 2000US-0225270P. | | |
| PR | 14-AUG-2000; 2000US-0225447P. | QY | 389 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQ 448 |
| PR | 14-AUG-2000; 2000US-0225477P. | Db | 1 GPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQ 60 |
| PR | 14-AUG-2000; 2000US-0225757P. | QY | 449 TRKEKDTNENRQKLRIFYQFLYNNTRQTEARDLHCPKCTLCNCRKLYSLKHLKLC 508 |
| PR | 22-AUG-2000; 2000US-0226888P. | Db | 61 TRKEKDTNENRQKLRIFYQFLYNNTRQTEARDLHCPKCTLCNCRKLYSLKHLKLC 120 |
| PR | 30-AUG-2000; 2000US-0228924P. | QY | 509 SRFIFNYVHPKGARIDVINECYDGSYAGNPQDIHQFGAFSRNGPVKRTPIITHLVC 568 |
| PR | 01-SEP-2000; 2000US-0228928P. | Db | 121 SRFIFNYVHPKGARIDVINEGYDGSYAGNPQDIHQFGAFSRNGPVKRTPIITHLVC 180 |
| PR | 01-SEP-2000; 2000US-0229343P. | QY | 569 RPRTKASMSFLESEGEVEQOQRTYSSGHNLRFHSTCLPLRPOEVEVSEDEKDEW 628 |
| PR | 01-SEP-2000; 2000US-0229344P. | Db | 181 RPRTKASMSFLESEGEVEQOQRTYSSGHNLRFHRTCLPLRPOEVEVSEDEKDEW 240 |
| PR | 05-SEP-2000; 2000US-0229345P. | QY | 629 LREKTIQIEEFSVDYNEGEKEVKMLNHLVMKGFADNQMNHACMLFVNYGQKIIKN 688 |
| PR | 05-SEP-2000; 2000US-0229503P. | Db | 241 LREKTIQIEEFSVDYNEGEKEVKMLNHLVMKGFADNQMNHACMLFVNYGQKIIKN 300 |
| PR | 21-SEP-2000; 2000US-0231413P. | QY | 689 LCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOKLEKGSASAPANEITEEQNGTANG 748 |
| PR | 21-SEP-2000; 2000US-0232232P. | Db | 301 LCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOKLEKGSASAPANEITEEQNGTANG 360 |
| PR | 25-SEP-2000; 2000US-0234274P. | QY | 749 FSEINSKEKALETDSVSGVSKOSKKQL 776 |
| PR | 27-SEP-2000; 2000US-0234997P. | Db | 361 FSEINSKEKALETDSVSGVSKOSKKQL 388 |
| PR | 29-SEP-2000; 2000US-0235632P. | | |
| PR | 29-SEP-2000; 2000US-0236367P. | | |
| PR | 29-SEP-2000; 2000US-0236368P. | | |
| PR | 29-SEP-2000; 2000US-0236369P. | | |
| PR | 29-SEP-2000; 2000US-0236370P. | | |
| PR | 02-OCT-2000; 2000US-0236802P. | | |
| PR | 02-OCT-2000; 2000US-0237037P. | | |
| PR | 02-OCT-2000; 2000US-0237038P. | | |
| PR | 02-OCT-2000; 2000US-0237039P. | | |
| PR | 02-OCT-2000; 2000US-0237040P. | | |
| PR | 13-OCT-2000; 2000US-0239935P. | | |
| PR | 20-OCT-2000; 2000US-0240960P. | | |
| PR | 20-OCT-2000; 2000US-0241785P. | | |
| PR | 20-OCT-2000; 2000US-0241809P. | | |
| PR | 01-NOV-2000; 2000US-0244617P. | | |
| PR | 17-NOV-2000; 2000US-0249299P. | | |
| PR | 08-DEC-2000; 2000US-0251856P. | | |
| PR | 08-DEC-2000; 2000US-0251868P. | | |
| PR | 08-DEC-2000; 2000US-0251869P. | | |
| PA | (ROSE/) ROSEN C A. | RESULT 8 | |
| PA | (RUBE/) RUBEN S M. | AAU16416 | |
| PA | (BARA/) BARASH S C. | ID AAU16416 standard; protein; 292 AA. | |
| XX | | XX AAU16416; | |
| XX | | XX | |
| PI | Rosen CA, Ruben SM, Barash SC; | 07-NOV-2001 (first entry) | |
| XX | | XX Human novel secreted protein, Seq ID 1369. | |
| XX | | XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; | |
| DR | WPI; 2003-147444/14. | XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; | |
| DR | N-PSDB; AEX73306. | XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; | |
| XX | | XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; | |
| PT | New polypeptides and nucleic acids, useful in gene therapy for treating, | XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; | |
| PT | inhibiting or preventing e.g. neural, immune system, muscular, | XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; | |
| PT | respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or | XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; | |
| PT | renal disorders. | XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; | |

XX PI Rosen CA, Barash SC, Ruben SM;
 XX DR WPI; 2001-488783/53.
 XX DR N-PSDB; AAS26403.
 XX PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX PS Claim 11; SEQ ID NO 1369; 980pp; English.
 XX CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Query Match 20.7%; Score 161; DB 4; Length 292;
 Best Local Similarity 100.0%; Pred. No. 1.5e-153;
 Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 354 VAMQEMECPISSKRAWETILDKRPPPTFGSGTLOFTLWTGETNDKSTAPIAKP 413
 DB 122 VAMQEMECPISSKRAWETILDKRPPPTFGSGTLOFTLWTGETNDKSTAPIAKP 181

OY 414 LATRNSLSLHGNKPGSVKPTQTAVKESLTDTQTRKEDTPNENKQKRIFFQFLYNN 473
 DB 182 LATRNSLSLHGNKPGSVKPTQTAVKESLTDTQTRKEDTPNENKQKRIFFQFLYNN 241

OY 474 NTRQQTAEADLHCPWCTLNCRKLYSLKHLKLCHSRPIFN 514
 DB 242 NTRQQTAEADLHCPWCTLNCRKLYSLKHLKLCHSRPIFN 282

RESULT 9
 ABUS5485
 ID ABUS5485 standard; protein; 292 AA.
 XX AC ABUS5485;
 XX DT 18-MAR-2003 (first entry)
 XX DE Human novel polypeptide #572.
 XX KW Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cycostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX OS Homo sapiens.

XX US2002132753-A1.
 XX 19-SEP-2002.
 XX PF 17-JAN-2001; 2001US-00764864.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 20-AUG-2000; 2000US-0226868P.
 PR 20-AUG-2000; 2000US-022824P.
 PR 01-SEP-2000; 2000US-0228287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 05-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 06-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2003-147444/14.
 DR N-PSDB; ABX73744.
 XX PT New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX Claim 11; SEQ ID NO 1369; 402pp; English.

XX The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC ABU55748 represent human novel polypeptides of the invention
XX
SQ Sequence 292 AA;
Query Match 20.7%; Score 161; DB 6; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.5e-153;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 354 VAMQWESCPISKKATWETILDGKRLPPFPFSGQPTLQTLRWGTNDKSTAPIAKP 413
Db 122 VAMQWESCPISKKATWETILDGKRLPPFPFSGQPTLQTLRWGTNDKSTAPIAKP 181
QY 414 LATRNSESLHQENKFGSVKPTQTIIVAKESLTTDLQTRKEKTPNENROKLRIFYQFLYNN 473
Db 182 LATRNSESLHQENKFGSVKPTQTIIVAKESLTTDLQTRKEKTPNENROKLRIFYQFLYNN 241
QY 474 NTRQTEARDLHCPWCTLNCRKLYSLKLKHLKCHSRFIFN 514
Db 242 NTRQTEARDLHCPWCTLNCRKLYSLKLKHLKCHSRFIFN 282
RESULT 10
AAU16402
ID AAU16402 standard; protein; 175 AA.
XX
AC AAU16402;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1355.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
FN WO200155322-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001341.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 11-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225677P.
PR 14-AUG-2000; 2000US-0225688P.
PR 14-AUG-2000; 2000US-0225707P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229433P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.

20-OCT-2000; 2000US-0241786P.
 20-OCT-2000; 2000US-0241787P.
 20-OCT-2000; 2000US-0241808P.
 20-OCT-2000; 2000US-0241809P.
 20-OCT-2000; 2000US-0241826P.
 01-NOV-2000; 2000US-0244617P.
 08-NOV-2000; 2000US-0246474P.
 08-NOV-2000; 2000US-0246475P.
 08-NOV-2000; 2000US-0246476P.
 08-NOV-2000; 2000US-0246477P.
 08-NOV-2000; 2000US-0246478P.
 08-NOV-2000; 2000US-0246523P.
 08-NOV-2000; 2000US-0246524P.
 08-NOV-2000; 2000US-0246525P.
 08-NOV-2000; 2000US-0246526P.
 08-NOV-2000; 2000US-0246527P.
 08-NOV-2000; 2000US-0246528P.
 08-NOV-2000; 2000US-0246532P.
 08-NOV-2000; 2000US-0246609P.
 08-NOV-2000; 2000US-0246610P.
 08-NOV-2000; 2000US-0246611P.
 08-NOV-2000; 2000US-0246613P.
 17-NOV-2000; 2000US-0249208P.
 17-NOV-2000; 2000US-0249209P.
 17-NOV-2000; 2000US-0249210P.
 17-NOV-2000; 2000US-0249211P.
 17-NOV-2000; 2000US-0249212P.
 17-NOV-2000; 2000US-0249213P.
 17-NOV-2000; 2000US-0249214P.
 17-NOV-2000; 2000US-0249215P.
 17-NOV-2000; 2000US-0249216P.
 17-NOV-2000; 2000US-0249217P.
 17-NOV-2000; 2000US-0249218P.
 17-NOV-2000; 2000US-0249244P.
 17-NOV-2000; 2000US-0249245P.
 17-NOV-2000; 2000US-0249264P.
 17-NOV-2000; 2000US-0249265P.
 17-NOV-2000; 2000US-0249297P.
 17-NOV-2000; 2000US-0249299P.
 17-NOV-2000; 2000US-0249300P.
 01-DEC-2000; 2000US-0250160P.
 01-DEC-2000; 2000US-0250391P.
 05-DEC-2000; 2000US-0251030P.
 05-DEC-2000; 2000US-0256719P.
 05-DEC-2000; 2000US-0251988P.
 06-DEC-2000; 2000US-0251479P.
 08-DEC-2000; 2000US-0251856P.
 08-DEC-2000; 2000US-0251869P.
 08-DEC-2000; 2000US-0251989P.
 08-DEC-2000; 2000US-0251990P.
 11-DEC-2000; 2000US-0254097P.
 05-JAN-2001; 2001US-0259678P.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-488783/53.
 N-PSDB; AAS26389.
 New nucleic acid molecules encoding 461 human secreted proteins for
 diagnosing, preventing, treating or ameliorating medical conditions and
 used as food additives or preservatives.
 Claim 11; SEQ ID NO 1355; 980pp; English.
 The invention relates to isolated nucleic acid molecules and their
 encoded secreted proteins. The nucleic acids and proteins are used to
 prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 in diagnosing a pathological condition or susceptibility to a

pathological condition. Antibodies to the proteins can also be used in
 alleviating symptoms associated with the disorders and in diagnostic
 immunosays e.g. radioimmunoassays or enzyme linked immunosorbant assays
 (ELISA). Disorders which are diagnosed or treated include autoimmune
 diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 arrest, cerebrovascular disorders e.g. cerebral ischaemia, angogenesis,
 nervous system disorders e.g. Alzheimer's disease, infections caused by
 bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 and many other disorders listed in the specification. The polypeptides
 can also be used to aid wound healing and epithelial cell proliferation,
 to prevent skin aging due to sunburn, to maintain organs before
 transplantation, for supporting cell culture of primary tissues, to
 regenerate tissues and in chemotaxis. The polypeptides can also be used
 as a food additive or preservative to increase or decrease storage
 capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 minerals, cofactors and other nutritional components. The present
 sequence represents a novel secreted protein of the invention. Note: The
 sequence data for this patent did not form part of the printed

Query Match 16.9%; Score 131; DB 4; Length 175;
 Best Local Similarity 100.0%; Pred. No. 1.9e-123;
 Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LYNNTRQQTTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIFNYVHPKGARDIVSIN 529
 |||||
 Db 5 LYNNTRQQTTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIFNYVHPKGARDIVSIN 64
 |||||

QY 530 ECVGSGYAGNPQDIHQPGFAFRNGPVKRTPTIHLVCRPKTKASMSFESEDEGEVE 589
 |||||
 Db 65 ECVGSGYAGNPQDIHQPGFAFRNGPVKRTPTIHLVCRPKTKASMSFESEDEGEVE 124
 |||||

QY 590 QORTYSSGHNH 600
 |||||
 Db 125 QORTYSSGHNH 135
 |||||

RESULT 11
 ABUS5471
 ID ABUS5471 standard; protein; 175 AA.
 AC ABUS5471;
 XX
 DT 18-MAR-2003 (first entry)
 XX
 DE Human novel polypeptide #558.
 XX
 KW Human; neural disorder; immune system disorder; renal disorder;
 muscular disorder; respiratory disease; reproductive disorder;
 gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 hyperproliferative disorder; inflammatory disease; allergic reaction;
 blood related disorder; cancer; immunosuppressive; antiinflammatory;
 cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 haemostatic; antiarteriosclerotic.
 XX
 OS Homo sapiens.
 XX
 XX US2002132753-A1.
 XX
 XX 19-SEP-2002.
 XX
 XX 17-JAN-2001; 2001US-00764864.
 XX
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214888P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 14-JUL-2000; 2000US-0217496P.
 PR 11-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR

CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumours). The cDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence presented is a protein of the invention. Note: Some of the
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.

SQ Sequence 243 AA;

Query Match 16.5%; Score 128; DB 7; Length 243;
 Best Local Similarity 100.0%; Pred. No. 2.8e-120; Indels 0; Gaps 0;
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFFSNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTYVALSY 60
 DB 1 MTGIAAASFFSNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTYVALSY 60

QY 61 INRFMTDAARREQESLKKKIOPKLSLTLSVSRGNVSTPPRHSSGSLTPVTPPTPSS 120
 DB 61 INRFMTDAARREQESLKKKIOPKLSLTLSVSRGNVSTPPRHSSGSLTPVTPPTPSS 120

QY 121 SFRSSTPT 128
 DB 121 SFRSSTPT 128

RESULT 13
 ADC06802
 ID ADC06802 standard; protein; 243 AA.
 XX AC ADC06802;
 XX DT 18-DEC-2003 (first entry)
 XX DE Human prostate cancer-related protein PCP0840.
 XX KW cytostatic; prostate cancer; breast; gene therapy; transgenic; human;
 XX KW chromosome 7p15.1.
 XX OS Homo sapiens.
 XX PN WO2003064599-A2.
 XX PD 07-AUG-2003.
 XX PF 24-JAN-2003; 2003WO-US001943.
 XX PR 25-JAN-2002; 2002US-00054935.
 XX PR 14-FEB-2002; 2002US-0356130P.
 XX PR 22-MAR-2002; 2002US-00102946.
 XX PR 08-APR-2002; 2002US-00117229.
 XX PR 14-MAY-2002; 2002US-00144198.
 XX PR 19-JUL-2002; 2002US-00197824.
 XX PA (ORIG-) ORIGENE TECHNOLOGIES INC.
 XX PI Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;
 XX DR WPI; 2003-679495/64.

XX New isolated polynucleotide related to cancer genes, useful for
 PT detecting, diagnosing, staging, monitoring, prognosticating, preventing
 PT or treating cancers, e.g. breast and prostate cancers.
 XX Claim 5; Fig 6; 128pp; English.
 XX The invention relates to a novel isolated polynucleotide comprising a

CC differentially-regulated mammalian cancer gene. The polynucleotides of
 CC the invention demonstrate cytostatic activity and are differentially
 CC expressed in prostate cancer. The polynucleotide, polypeptides and
 CC methods of the invention may be useful for detecting, diagnosing,
 CC staging, monitoring, prognosticating, preventing or treating cancers,
 CC particularly breast and prostate cancers. Furthermore, the invention may
 CC be utilised during gene therapy procedures or in the production of
 CC transgenic animals. The current sequence is that of the prostate cancer-
 CC related protein of the invention.

SQ Sequence 243 AA;

Query Match 16.5%; Score 128; DB 7; Length 243;
 Best Local Similarity 100.0%; Pred. No. 2.8e-120; Indels 0; Gaps 0;
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFFSNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTYVALSY 60
 DB 1 MTGIAAASFFSNTCRFGCGGLHFTPLADLIEHIEDNHIDTPRVLEKQELQOPTYVALSY 60

QY 61 INRFMTDAARREQESLKKKIOPKLSLTLSVSRGNVSTPPRHSSGSLTPVTPPTPSS 120
 DB 61 INRFMTDAARREQESLKKKIOPKLSLTLSVSRGNVSTPPRHSSGSLTPVTPPTPSS 120

QY 121 SFRSSTPT 128
 DB 121 SFRSSTPT 128

RESULT 14
 AAU16381
 ID AAU16381 standard; protein; 278 AA.
 XX AC AAU16381;
 XX DT 07-NOV-2001 (first entry)
 XX DE Human novel secreted protein, Seq ID 1334.
 XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
 KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
 KW antibacterial; virucide; fungicide; ophthalmological; vulnery;
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;
 KW preservative; antiproliferative.

XX OS Homo sapiens.
 XX PN WO200155322-A2.
 XX PD 02-AUG-2001.
 XX PF 17-JAN-2001; 2001WO-US001341.
 XX PR 31-JAN-2000; 2000US-0179065P.
 XX PR 04-FEB-2000; 2000US-0180628P.
 XX PR 24-FEB-2000; 2000US-0184664P.
 XX PR 02-MAR-2000; 2000US-0186350P.
 XX PR 16-MAR-2000; 2000US-0189874P.
 XX PR 17-MAR-2000; 2000US-0190076P.
 XX PR 18-APR-2000; 2000US-0198123P.
 XX PR 19-MAY-2000; 2000US-0205515P.
 XX PR 07-JUN-2000; 2000US-0209467P.
 XX PR 28-JUN-2000; 2000US-0214886P.
 XX PR 30-JUN-2000; 2000US-021513P.
 XX PR 07-JUL-2000; 2000US-0216647P.
 XX PR 07-JUL-2000; 2000US-0216880P.
 XX PR 11-JUL-2000; 2000US-0217487P.
 XX PR 11-JUL-2000; 2000US-0217496P.
 XX PR 14-JUL-2000; 2000US-0218290P.

| | | | | | |
|----|--------------|------------------|----|---|------------------|
| PR | 26-JUL-2000; | 2000US-0220963P. | PR | 08-NOV-2000; | 2000US-0246475P. |
| PR | 26-JUL-2000; | 2000US-0220964P. | PR | 08-NOV-2000; | 2000US-0246476P. |
| PR | 14-AUG-2000; | 2000US-0224518P. | PR | 08-NOV-2000; | 2000US-0246477P. |
| PR | 14-AUG-2000; | 2000US-0224519P. | PR | 08-NOV-2000; | 2000US-0246478P. |
| PR | 14-AUG-2000; | 2000US-0225213P. | PR | 08-NOV-2000; | 2000US-0246523P. |
| PR | 14-AUG-2000; | 2000US-0225214P. | PR | 08-NOV-2000; | 2000US-0246524P. |
| PR | 14-AUG-2000; | 2000US-0225266P. | PR | 08-NOV-2000; | 2000US-0246525P. |
| PR | 14-AUG-2000; | 2000US-0225267P. | PR | 08-NOV-2000; | 2000US-0246526P. |
| PR | 14-AUG-2000; | 2000US-0225268P. | PR | 08-NOV-2000; | 2000US-0246527P. |
| PR | 14-AUG-2000; | 2000US-0225270P. | PR | 08-NOV-2000; | 2000US-0246528P. |
| PR | 14-AUG-2000; | 2000US-0225447P. | PR | 08-NOV-2000; | 2000US-0246532P. |
| PR | 14-AUG-2000; | 2000US-0225757P. | PR | 08-NOV-2000; | 2000US-0246609P. |
| PR | 14-AUG-2000; | 2000US-0225758P. | PR | 08-NOV-2000; | 2000US-0246610P. |
| PR | 14-AUG-2000; | 2000US-0225759P. | PR | 08-NOV-2000; | 2000US-0246611P. |
| PR | 18-AUG-2000; | 2000US-0226279P. | PR | 08-NOV-2000; | 2000US-0246613P. |
| PR | 22-AUG-2000; | 2000US-0226681P. | PR | 17-NOV-2000; | 2000US-0249207P. |
| PR | 22-AUG-2000; | 2000US-0226686P. | PR | 17-NOV-2000; | 2000US-0249208P. |
| PR | 22-AUG-2000; | 2000US-0227182P. | PR | 17-NOV-2000; | 2000US-0249209P. |
| PR | 23-AUG-2000; | 2000US-0227009P. | PR | 17-NOV-2000; | 2000US-0249210P. |
| PR | 30-AUG-2000; | 2000US-0228924P. | PR | 17-NOV-2000; | 2000US-0249211P. |
| PR | 01-SEP-2000; | 2000US-0228287P. | PR | 17-NOV-2000; | 2000US-0249212P. |
| PR | 01-SEP-2000; | 2000US-0229343P. | PR | 17-NOV-2000; | 2000US-0249213P. |
| PR | 01-SEP-2000; | 2000US-0229344P. | PR | 17-NOV-2000; | 2000US-0249214P. |
| PR | 01-SEP-2000; | 2000US-0229345P. | PR | 17-NOV-2000; | 2000US-0249215P. |
| PR | 05-SEP-2000; | 2000US-0229503P. | PR | 17-NOV-2000; | 2000US-0249216P. |
| PR | 05-SEP-2000; | 2000US-0229513P. | PR | 17-NOV-2000; | 2000US-0249217P. |
| PR | 06-SEP-2000; | 2000US-0230437P. | PR | 17-NOV-2000; | 2000US-0249218P. |
| PR | 06-SEP-2000; | 2000US-0230438P. | PR | 17-NOV-2000; | 2000US-0249219P. |
| PR | 08-SEP-2000; | 2000US-0231242P. | PR | 17-NOV-2000; | 2000US-0249244P. |
| PR | 08-SEP-2000; | 2000US-0231243P. | PR | 17-NOV-2000; | 2000US-0249245P. |
| PR | 08-SEP-2000; | 2000US-0231244P. | PR | 17-NOV-2000; | 2000US-0249246P. |
| PR | 08-SEP-2000; | 2000US-0231413P. | PR | 17-NOV-2000; | 2000US-0249265P. |
| PR | 08-SEP-2000; | 2000US-0231414P. | PR | 17-NOV-2000; | 2000US-0249297P. |
| PR | 08-SEP-2000; | 2000US-0232080P. | PR | 17-NOV-2000; | 2000US-0249299P. |
| PR | 08-SEP-2000; | 2000US-0232081P. | PR | 17-NOV-2000; | 2000US-0249300P. |
| PR | 12-SEP-2000; | 2000US-0231969P. | PR | 01-DEC-2000; | 2000US-0250160P. |
| PR | 14-SEP-2000; | 2000US-0232397P. | PR | 01-DEC-2000; | 2000US-0250391P. |
| PR | 14-SEP-2000; | 2000US-0232398P. | PR | 05-DEC-2000; | 2000US-0251030P. |
| PR | 14-SEP-2000; | 2000US-0232399P. | PR | 05-DEC-2000; | 2000US-0251988P. |
| PR | 14-SEP-2000; | 2000US-0232400P. | PR | 06-DEC-2000; | 2000US-0256719P. |
| PR | 14-SEP-2000; | 2000US-0232401P. | PR | 06-DEC-2000; | 2000US-0251479P. |
| PR | 14-SEP-2000; | 2000US-0233063P. | PR | 08-DEC-2000; | 2000US-0251856P. |
| PR | 14-SEP-2000; | 2000US-0233064P. | PR | 08-DEC-2000; | 2000US-0251868P. |
| PR | 14-SEP-2000; | 2000US-0233065P. | PR | 08-DEC-2000; | 2000US-0251869P. |
| PR | 21-SEP-2000; | 2000US-0234223P. | PR | 08-DEC-2000; | 2000US-0251989P. |
| PR | 21-SEP-2000; | 2000US-0234274P. | PR | 08-DEC-2000; | 2000US-0251990P. |
| PR | 25-SEP-2000; | 2000US-0234997P. | PR | 11-DEC-2000; | 2000US-0254097P. |
| PR | 25-SEP-2000; | 2000US-0234998P. | PR | 05-JAN-2001; | 2001US-0259678P. |
| PR | 26-SEP-2000; | 2000US-0235484P. | XX | (HUMA-) HUMAN GENOME SCI INC. | |
| PR | 27-SEP-2000; | 2000US-0235834P. | PA | Rosen CA, Barash SC, Ruben SM; | |
| PR | 27-SEP-2000; | 2000US-0235836P. | XX | WPI; 2001-488783/53. | |
| PR | 29-SEP-2000; | 2000US-0236327P. | XX | N-PSDB; AAS26368. | |
| PR | 29-SEP-2000; | 2000US-0236367P. | DR | New nucleic acid molecules encoding 461 human secreted proteins for | |
| PR | 29-SEP-2000; | 2000US-0236368P. | DR | diagnosing, preventing, treating or ameliorating medical conditions and | |
| PR | 29-SEP-2000; | 2000US-0236369P. | XX | used as food additives or preservatives. | |
| PR | 29-SEP-2000; | 2000US-0236399P. | XX | Claim 11; SEQ ID NO 1334; 980pp; English. | |
| PR | 29-SEP-2000; | 2000US-0236399P. | PS | The invention relates to isolated nucleic acid molecules and their | |
| PR | 29-SEP-2000; | 2000US-0236370P. | CC | encoded secreted proteins. The nucleic acids and proteins are used to | |
| PR | 29-SEP-2000; | 2000US-0236370P. | CC | prevent, treat or ameliorate a medical condition in e.g. humans, mice, | |
| PR | 29-SEP-2000; | 2000US-0236399P. | CC | rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used | |
| PR | 02-OCT-2000; | 2000US-0236802P. | CC | in diagnosing a pathological condition or susceptibility to a | |
| PR | 02-OCT-2000; | 2000US-0237037P. | CC | pathological condition. Antibodies to the proteins can also be used in | |
| PR | 02-OCT-2000; | 2000US-0237038P. | CC | alleviating symptoms associated with the disorders and in diagnostic | |
| PR | 02-OCT-2000; | 2000US-0237039P. | CC | immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays | |
| PR | 02-OCT-2000; | 2000US-0237040P. | CC | (ELISA). Disorders which are diagnosed or treated include autoimmune | |
| PR | 13-OCT-2000; | 2000US-0239935P. | CC | diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. | |
| PR | 13-OCT-2000; | 2000US-0239937P. | CC | neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac | |
| PR | 20-OCT-2000; | 2000US-0240960P. | CC | arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, | |
| PR | 20-OCT-2000; | 2000US-0241221P. | | | |
| PR | 20-OCT-2000; | 2000US-0241785P. | | | |
| PR | 20-OCT-2000; | 2000US-0241787P. | | | |
| PR | 20-OCT-2000; | 2000US-0241808P. | | | |
| PR | 20-OCT-2000; | 2000US-0241809P. | | | |
| PR | 20-OCT-2000; | 2000US-0241836P. | | | |
| PR | 01-NOV-2000; | 2000US-0244617P. | | | |
| PR | 08-NOV-2000; | 2000US-0246474P. | | | |

CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Query Match 14.0%; Score 109; DB 4; Length 278;
 Best Local Similarity 100.0%; Pred. No. 5.1e-101; Indels 0; Gaps 0;
 Matches 109; Conservative 0; Mismatches 0;

QY 1 MTGIAAASFFNTCRFGCGGLHFTLADLIEHIEDNHIDTPRVLEKQELQQPTVALSY 60
 DB 36 MTGIAAASFFNTCRFGCGGLHFTLADLIEHIEDNHIDTPRVLEKQELQQPTVALSY 95
 QY 61 INRFMTDAARREQESLKKIKQPKLSLTSSSVSRGNVSTPPRHSSGSLT 109
 DB 96 INRFMTDAARREQESLKKIKQPKLSLTSSSVSRGNVSTPPRHSSGSLT 144

RESULT 15
 ABUS5450
 ID ABUS5450 standard; protein; 278 AA.

AC ABUS5450;
 DT 18-MAR-2003 (first entry)
 XX Human novel polypeptide #537.
 DE
 XX Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.

OS Homo sapiens.

PN US2002132753-A1.

XX 19-SEP-2002.

PF 17-JAN-2001; 2001US-00764864.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 28-JUN-2000; 2000US-0214886P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 14-JUL-2000; 2000US-0217496P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 22-AUG-2000; 2000US-0226868P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239353P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 XX WPI; 2003-147444/14.
 DR N-PSDB; ABX73709.

XX New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX
 PS Claim 11; SEQ ID NO 1334; 402pp; English.

XX The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
 CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and
 CC ABUS5748 represent human novel polypeptides of the invention

XX Sequence 278 AA;

Query Match 14.0%; Score 109; DB 6; Length 278;

Best Local Similarity 100.0%; Pred. No. 5.1e-101; Indels 0; Gaps 0;

Matches 109; Conservative 0; Mismatches 0;

QY 1 MTGIAAASFFNTCRFGCGGLHFTLADLIEHIEDNHIDTPRVLEKQELQQPTVALSY 60

DB 36 MTGIAAASFFNTCRFGCGGLHFTLADLIEHIEDNHIDTPRVLEKQELQQPTVALSY 95

QY 61 INRFMTDAARREQESLKKIKQPKLSLTSSSVSRGNVSTPPRHSSGSLT 109

Db 96 INRFMTDARRRQESLKKIQPKLSLTLSSSVSRGCVSTPPRHSSGSLT 144

Search completed: August 25, 2004, 18:11:48
Job time : 136 secs

1000

1000

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:25:28 ; Search time 132 Seconds
(without alignments)
1661.037 Million cell updates/sec

Title: US-09-874-162A-8

Perfect score: 4078
Sequence: 1 MTGIAAASFNTCRFGGCG.....KALETDSVSGVSKQXKQL 776

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|---------------------|
| 1 | 3725 | 91.3 | 776 | 5 | Abg30843 Human JAZ |
| 2 | 3413 | 83.7 | 803 | 3 | Abb00066 KIAA0160 |
| 3 | 3237 | 79.4 | 739 | 5 | Abc30842 Human Jai |
| 4 | 2043 | 50.1 | 388 | 4 | Aau15978 Human nov |
| 5 | 2043 | 50.1 | 388 | 6 | Abu55047 Human nov |
| 6 | 1497 | 36.7 | 289 | 4 | Aau15958 Human nov |
| 7 | 1497 | 36.7 | 289 | 6 | Abu55027 Human nov |
| 8 | 1453 | 35.6 | 292 | 4 | Aau16416 Human nov |
| 9 | 1453 | 35.6 | 292 | 6 | Abu55485 Human nov |
| 10 | 1127.5 | 27.6 | 955 | 4 | Abb71635 Drosophil |
| 11 | 919 | 22.5 | 175 | 4 | Aau16402 Human nov |
| 12 | 919 | 22.5 | 175 | 6 | Abu55471 Human nov |
| 13 | 661 | 16.2 | 243 | 7 | Adc064250 Human pro |
| 14 | 661 | 16.2 | 243 | 7 | Adc06802 Human pro |
| 15 | 652 | 16.0 | 278 | 4 | Aau16381 Human nov |
| 16 | 652 | 16.0 | 278 | 6 | Abu55450 Human nov |
| 17 | 490 | 12.0 | 243 | 5 | Abg30841 Human jux |
| 18 | 476 | 11.7 | 143 | 4 | Aau15931 Human nov |
| 19 | 476 | 11.7 | 143 | 6 | Abu55000 Human nov |
| 20 | 437.5 | 10.7 | 102 | 3 | Aag03182 Human sec |
| 21 | 429 | 10.5 | 202 | 3 | Aab43387 Human ORF |
| 22 | 280 | 6.9 | 171 | 4 | Abb11690 Human SFP |
| 23 | 238.5 | 5.8 | 604 | 3 | Aay53933 An Os-MPC |
| 24 | 230 | 5.6 | 498 | 3 | Aag54168 Arabidops |
| 25 | 227.5 | 5.6 | 367 | 3 | Aag54169 Arabidops |

| | | | | | | |
|----|-------|-----|------|---|----------|--------------------|
| 26 | 224 | 5.5 | 611 | 3 | AAY53932 | AAY53932 A MPC1 pr |
| 27 | 224 | 5.5 | 767 | 4 | ABB58240 | ABB58240 Drosophil |
| 28 | 221 | 5.4 | 445 | 3 | AAB00060 | AAB00060 VRN2 poly |
| 29 | 217 | 5.3 | 440 | 3 | AAB00061 | AAB00061 VRN2 poly |
| 30 | 211 | 5.2 | 692 | 6 | ABP70524 | ABP70524 Histone d |
| 31 | 211 | 5.2 | 813 | 3 | AAB01674 | AAB01674 F1s2 prot |
| 32 | 176.5 | 4.3 | 1142 | 4 | AAG70713 | AAG70713 S cerevis |
| 33 | 176.5 | 4.3 | 1142 | 6 | ABR53123 | ABR53123 Protein s |
| 34 | 176.5 | 4.3 | 1349 | 4 | AAG70854 | AAG70854 C albica |
| 35 | 161.5 | 4.0 | 583 | 5 | ABG93227 | ABG93227 C. albica |
| 36 | 159.5 | 3.9 | 339 | 3 | AAG54170 | AAG54170 Arabidops |
| 37 | 156.5 | 3.8 | 3696 | 5 | ABP40235 | ABP40235 Staphyloc |
| 38 | 153.5 | 3.8 | 1279 | 4 | AAG83047 | AAG83047 S. epider |
| 39 | 153.5 | 3.8 | 3692 | 6 | ABU43311 | ABU43311 Protein e |
| 40 | 151 | 3.7 | 1881 | 5 | ABP73809 | ABP73809 Candida a |
| 41 | 150 | 3.7 | 912 | 2 | AAR36731 | AAR36731 Ubiquitin |
| 42 | 149 | 3.7 | 1087 | 2 | AAY19935 | AAY19935 B. burgdo |
| 43 | 149 | 3.7 | 1119 | 2 | AAY19934 | AAY19934 B. burgdo |
| 44 | 147.5 | 3.6 | 1790 | 6 | ABR53116 | ABR53116 Protein s |
| 45 | 147 | 3.6 | 907 | 4 | ABU53073 | ABU53073 Intracell |

ALIGNMENTS

RESULT 1
ABG30843
ID ABG30843 standard; protein; 776 AA.

XX AC ABG30843;

DT 21-OCT-2002 (first entry)

DE Human JAZF1/jJAZ1 fusion protein.

KW Human; JAZF1; juxtaposed with another zinc finger; jJAZ1; JAZF1/jJAZ1;
joined with JAZF1; proliferation; endometrial stroma tumour; immunogen;
antigen; antibody; fertility; pregnancy; gene therapy; vaccine;
fusion protein.

XX OS Homo sapiens.

OS Synthetic.

OS Chimeric.

XX Key Location/Qualifiers

FT Misc-difference 13..37
/note= "Encoded by ACCTGCCGATTCGGGGCTGGGGACTCCACTTC-
CCACCTCGCCGACCTCATCGAGCACATGAGGACACAC. C2H2 zinc
finger domain"

FT Region
/note= "Breakpoint for production of fusion protein
JAZF1/jJAZ1"

FT Misc-difference 357
/note= "Encoded by CAG"

FT Misc-difference 485..508
/note= "Encoded by CTGCATTGCCCTTGGTGTACTCTGAAGTGGCGAA-
ACCTTATAGTTACTCAAGCATCTTAAGCTCTGCCAT. C2H2 zinc finger
domain"

FT Region
/note= "Bipartite nuclear localisation signal"

XX WO200193805-A2.

XX 13-DEC-2001.

XX 04-JUN-2001; 2001WO-US017936.

XX 02-JUN-2000; 2000US-0209093P.

XX (BGHM) BRIGHAM & WOMENS HOSPITAL INC.

XX Koontz J, Sklar J;

DR WPI; 2002-575047/61.
DR N-PSDB; ABR89163.
XX Novel JAZF1, jJAZ1 or JAZF1/jJAZ1 polypeptides useful as immunogens or
PT antigens to raise or test anti-JAZF1, jJAZ1 or JAZF1/jJAZ1 antibodies.
XX
PS Claim 7; fig 3; 76pp; English.
XX
CC The present invention relates to a new JAZF1 (juxtaposed with another
CC zinc finger), jJAZ1 (joined with JAZF1) or JAZF1/jJAZ1 polypeptide. The
CC methods of the invention can be used to identify a compound which
CC controls proliferation of endometrial stroma, by expressing jJAZ1 in the
CC presence of the compound, and determining whether the compound affects
CC expression of jJAZ1. JAZF1, jJAZ1 or JAZF1/jJAZ1 polypeptides are useful
CC as immunogens or antigens to raise or test anti-JAZF1, jJAZ1 or
CC JAZF1/jJAZ1 antibodies. The invention can be used as bait proteins in a
CC two hybrid assay or three hybrid assay to identify other proteins which
CC bind or interact with JAZF1/jJAZ1-binding proteins. JAZF1, jJAZ1 or
CC JAZF1/jJAZ1 molecules are useful for identifying the origin of tumour and
CC as tumour marker protein to verify that a stromal tumour is from
CC endometrium. The antibody is useful for promoting or decreasing fertility
CC or pregnancy, and also for treating endometrial stromal tumours. The
CC present amino acid sequence represents the human JAZF1/jJAZ1 fusion
CC protein of the invention
XX
SQ Sequence 776 AA;
Query Match 91.3%; Score 3725; DB 5; Length 776;
Best Local Similarity 93.6%; Pred. No. 1.4e-311;
Matches 726; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 1 MTGTAASFFSNTCFGCGGLHFTPLADLIEHDNHTDTPRVLEKQELQOQPTVVALSY 60
DB 1 MTGTAASFFSNTCFGCGGLHFTPLADLIEHDNHTDTPRVLEKQELQOQPTVVALSY 60
QY 61 INRFMTDAARQESLKKKIQKLSLTSSSVRGVSTPPRHSSGSLTPVTPPTPSS 120
DB 61 INRFMTDAARQESLKKKIQKLSLTSSSVRGVSTPPRHSSGSLTPVTPPTPSS 120
QY 121 SFRSSTPEPTQIVRFLRNLIAPIFLHRTLTYMHSRNTNKRKTFKVDMLSKVEK 180
DB 121 SFRSSTPEPTQIVRFLRNLIAPIFLHRTLTYMHSRNTNKRKTFKVDMLSKVEK 180
QY 181 MKGEQESHLSAHLQLTFTGFHNDKPSNSEQNSVTLVLLVKVCHKKRDXVSCPI 240
DB 181 MKGEQESHLSAHLQLTFTGFHNDKPSNSEQNSVTLVLLVKVCHKKRDXVSCPI 240
QY 241 RQVPTGKQVPLIDPLNQTGKGNFSLAVSNEFEPNSHMVKSYSLLFRVTRGREFN 300
DB 241 RQVPTGKQVPLIDPLNQTGKGNFSLAVSNEFEPNSHMVKSYSLLFRVTRGREFN 300
QY 301 GMINGETNENIDVNEELPARKRNREDGEKTFVAQMTVFDKQRLQLLDGEYEVAMQEME 360
DB 301 GMINGETNENIDVNEELPARKRNREDGEKTFVAQMTVFDKQRLQLLDGEYEVAMQEME 360
QY 361 ECPISKKEATWETLDGRLPPPTFTFSGQPTLQTLRWGTGNTDKSTAPIAKPLATNSE 420
DB 361 ECPISKKEATWETLDGRLPPPTFTFSGQPTLQTLRWGTGNTDKSTAPIAKPLATNSE 420
QY 421 SLHQSNKPGSVKPTQTIKVESLTDLTQTRKEDTPNENKQKLRIFYQFLYNNNTROTE 480
DB 421 SLHQSNKPGSVKPTQTIKVESLTDLTQTRKEDTPNENKQKLRIFYQFLYNNNTROTE 480
QY 481 ARDDLHCPWCTNCRKLYSLKHLKCHSRFIFNYVHPKGRIDVSNICYDGSYAGNP 540
DB 481 ARDDLHCPWCTNCRKLYSLKHLKCHSRFIFNYVHPKGRIDVSNICYDGSYAGNP 540
QY 541 QDIHQPGFASRNGPVKRTPTIHLVCRPRTKASMSFELESDEGEVQORTYSSGHR 600
DB 541 QDIHQPGFASRNGPVKRTPTIHLVCRPRTKASMSFELESDEGEVQORTYSSGHR 600
QY 601 LYFHSDDTCLPLRPOEMEVDSDEKDPWLRKTTITQIEEFSVDNVEGEKVMKLNHLVMK 660

DB 601 LYFHSDDTCLPLRPOEMEVDSDEKDPWLRKTTITQIEEFSVDNVEGEKVMKLNHLVMK 660
QY 661 HGFIADQNQNHACMLFVENYQKIIKKNLCRNFMHLVSMHDFNLISIMSDKAVTKLRE 720
DB 661 HGFIADQNQNHACMLFVENYQKIIKKNLCRNFMHLVSMHDFNLISIMSDKAVTKLRE 720
QY 721 MQOKLEKGSASAPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776
DB 721 MQOKLEKGSASAPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776
RESULT 2
AAB00066
ID AAB00066 standard; protein; 803 AA.
XX AAB00066;
XX 16-NOV-2000 (first entry)
XX KIAA0160 polypeptide.
XX Vernalization gene; VRN2; plant characteristic; flowering time;
XX leaf size; leaf shape; shade avoidance response; reproduction; breeding;
XX KW pollination; cultivation; human.
XX OS Homo sapiens.
XX PN WO200044918-A1.
XX PD 03-AUG-2000.
XX 28-JAN-2000; 2000WO-GB000248.
XX 28-JAN-1999; 99GB-00001927.
XX (PLAN-) PLANT BIOSCIENCE LTD.
XX Dean C, Gendall A;
XX WPI; 2000-499333/44.
XX N-PSDB; AAA47759.
XX Isolated vernalization gene VRN2 is used to produce transgenic plants
XX with altered vernalization response, flowering time, leaf size and/or
XX shape or shade avoidance response for maximized reproductive success.
XX Disclosure; Page 79; 105pp; English.
XX Isolated nucleic acid sequences obtained from the VRN2 locus of a plant
XX encode polypeptides which are capable of affecting one or more
XX vernalization responses such as, flowering time, leaf size and/or shape
XX or the shade avoidance response of a plant into which the nucleic acid is
XX introduced. Introducing such sequences into plants to alter these
XX characteristics maximises the reproductive success of the plant. This
XX protein is encoded an isolated human sequence which has homology to the
XX Arabidopsis thaliana VRN2 gene over a short region near the N-terminus
SQ Sequence 803 AA;
Query Match 83.7%; Score 3413; DB 3; Length 803;
Best Local Similarity 99.8%; Pred. No. 1.2e-284;
Matches 647; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 129 EPTQYRFLRNLIAPIFLHRTLTYMHSRNTNKRKTFKVDMLSKVEMKGEQESH 188
DB 129 EPTQYRFLRNLIAPIFLHRTLTYMHSRNTNKRKTFKVDMLSKVEMKGEQESH 188
QY 189 SLSAHLQLTFTGFHNDKPSNSEQNSVTLVLLVKVCHKKRDXVSCPIRQVPTGK 248
DB 189 SLSAHLQLTFTGFHNDKPSNSEQNSVTLVLLVKVCHKKRDXVSCPIRQVPTGK 248
QY 216 SUSAHLQLTFTGFHNDKPSNSEQNSVTLVLLVKVCHKKRDXVSCPIRQVPTGK 275
DB 216 SUSAHLQLTFTGFHNDKPSNSEQNSVTLVLLVKVCHKKRDXVSCPIRQVPTGK 275
QY 249 QVPLIPDLNQTGKGNFSLAVSNEFEPNSHMVKSYSLLFRVTRGREFNMGNTN 308
DB 249 QVPLIPDLNQTGKGNFSLAVSNEFEPNSHMVKSYSLLFRVTRGREFNMGNTN 308

Db 276 QVPLIPDLNQTGKGNFSLAVSSNEFEPNSHMVKSYSLLFRVTRPGRRREFNGMGETN 335
 Qy 309 ENIDVNEELPARKRNEDGKTFVAQMTVFDKRRRLQLLDGEYVAMQEMEECPISKRR 368
 Db 336 ENIDVNEELPARKRNEDGKTFVAQMTVFDKRRRLQLLDGEYVAMQEMEECPISKRR 395
 Qy 369 ATWETILDGKRLPPPTETFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKP 428
 Db 396 ATWETILDGKRLPPPTETFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKP 455
 Qy 429 GSVKPTQTIAVKESLTDLQTRKEKDTNENRQKLRIFYQFLYNNTRQOTEARDLHCP 488
 Db 456 GSVKPTQTIAVKESLTDLQTRKEKDTNENRQKLRIFYQFLYNNTRQOTEARDLHCP 515
 Qy 489 WCTLNCRKLYSLKHLKCHSRPIFYNYVHPKGAIDVINECYDGSVAGNPQDIHQPG 548
 Db 516 WCTLNCRKLYSLKHLKCHSRPIFYNYVHPKGAIDVINECYDGSVAGNPQDIHQPG 575
 Qy 549 FAFSRNGPVKRTPTIHLVCRPKRTKASMBFLESEDEGEVQQRYSGGHNLRYFHSDDTC 608
 Db 576 FAFSRNGPVKRTPTIHLVCRPKRTKASMBFLESEDEGEVQQRYSGGHNLRYFHSDDTC 635
 Qy 609 LPLRPQEMEVDSEDEKDPWLREKTTITQIEEFSVDNVEGEKVMKLNHLVHMKGHFIADNQ 668
 Db 636 LPLRPQEMEVDSEDEKDPWLREKTTITQIEEFSVDNVEGEKVMKLNHLVHMKGHFIADNQ 695
 Qy 669 MNHACMLFVENYQCKIIKKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOQKLEKG 728
 Db 696 MNHACMLFVENYQCKIIKKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOQKLEKG 755
 Qy 729 ESASPANBEITTEQNGTANGSEINSEKSALETDSVSGVSKQSKQKL 776
 Db 756 ESASPANBEITTEQNGTANGSEINSEKSALETDSVSGVSKQSKQKL 803

RESULT 3

ID ABG30842 standard; protein; 739 AA.
 AC ABG30842;

DT 21-OCT-2002 (first entry)
 XX Human joined with JAZF1 (JJAZ1) protein.

DE Human; JAZF1; juxtaposed with another zinc finger; JJAZ1; JAZF1/JJAZ1;
 KW joined with JAZF1; proliferation; endometrial stroma tumour; immunogen;
 KW antigen; antibody; fertility; pregnancy; gene therapy; vaccine;
 KW chromosome 17.

OS Homo sapiens.

FH Key Location/Qualifiers
 FT Region 92..93

FT /note= "Breakpoint for production of fusion protein
 JAZF1/JJAZ1"

FT Misc-difference 448..471

FT /note= "Encoded by CNGCATGCCCTGTGTACTGTGAACCTGCCGCA-
 ACTTATAGTTTACTCAAGCATCTAAACTCTGCAT. C2H2 zinc finger
 domain"

FT Region 521..538

FT /note= "Bipartite nuclear localisation signal"

XX WO200193805-A2.

XX 13-DEC-2001.

XX 04-JUN-2001; 2001WO-US017936.

XX 02-JUN-2000; 2000US-0209093P.

XX (BGHM) BRIGHAM & WOMENS HOSPITAL INC.

PA

PI

Koontz J, Sklar J;

DR WPI; 2002-575047/61.

XX N-PSDB; ABK89162.

PT Novel JAZF1, JJAZ1 or JAZF1/JJAZ1 polypeptides useful as immunogens or
 antigens to raise or test anti-JAZF1, JJAZ1 or JAZF1/JJAZ1 antibodies.

PS Claim 5; Fig 2; 76pp; English.

XX The present invention relates to a new JAZF1 (juxtaposed with another
 zinc finger), JJAZ1 (joined with JAZF1) or JAZF1/JJAZ1 polypeptide. The
 methods of the invention can be used to identify a compound which
 controls proliferation of endometrial stroma, by expressing JJAZ1 in the
 presence of the compound, and determining whether the compound affects
 expression of JJAZ1, JJAZ1 or JAZF1/JJAZ1 polypeptides are useful
 as immunogens or antigens to raise or test anti-JAZF1, JJAZ1 or
 JAZF1/JJAZ1 antibodies. The invention can be used as bait proteins in a
 two hybrid assay or three hybrid assay to identify other proteins which
 bind or interact with JAZF1/JJAZ1-binding proteins. JAZF1, JJAZ1 or
 JAZF1/JJAZ1 molecules are useful for identifying the origin of tumour and
 as tumour marker protein to verify that a stromal tumour is from
 endometrium. The antibody is useful for promoting or decreasing fertility
 or pregnancy, and also for treating endometrial stromal tumours. The
 present amino acid sequence represents the human JJAZ1 protein of the
 invention. This sequence is encoded by the human JJAZ1 gene located on
 chromosome 17

SQ Sequence 739 AA;

Query Match 79.4%; Score 3237; DB 5; Length 739;

Best Local Similarity 96.1%; Pred. No. 1.5e-269;

Matches 623; Conservative 1; Mismatches 24; Indels 0; Gaps 0;

Qy 129 EPTQIVRFLRTRNLIAPIFLHRTLTYMSHRNSTNKRKTFKVDMLSKVKMKQGQESH 188
 Db 92 KPTQIVRFLRTRNLIAPIFLHRTLTYMSHRNSTNKRKTFKVDMLSKVKMKQGQESH 151
 Qy 189 SLSAHLQLTFTGTFHKNKDPSPNSNEQNSVTLVLLVVKVCHKKKDVSCPIRQVPTGKK 248
 Db 152 SLSAHLQLTFTGTFHKNKDPSPNSNEQNSVTLVLLVVKVCHKKKDVSCPIRQVPTGKK 211
 Qy 249 QVPLIPDLNQTGKGNFSLAVSSNEFEPNSHMVKSYSLLFRVTRPGRRREFNGMGETN 308
 Db 212 QVPLIPDLNQTGKGNFSLAVSSNEFEPNSHMVKSYSLLFRVTRPGRRREFNGMGETN 271
 Qy 309 ENIDVNEELPARKRNEDGKTFVAQMTVFDKRRRLQLLDGEYVAMQEMEECPISKRR 368
 Db 272 ENIDVNEELPARKRNEDGKTFVAQMTVFDKRRRLQLLDGEYVAMQEMEECPISKRR 331
 Qy 369 ATWETILDGKRLPPPTETFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKP 428
 Db 332 ATWETILDGKRLPPPTETFSQPTLQFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKP 391
 Qy 429 GSVKPTQTIAVKESLTDLQTRKEKDTNENRQKLRIFYQFLYNNTRQOTEARDLHCP 488
 Db 392 GSVKPTQTIAVKESLTDLQTRKEKDTNENRQKLRIFYQFLYNNTRQOTEARDLHCP 451
 Qy 489 WCTLNCRKLYSLKHLKCHSRPIFYNYVHPKGAIDVINECYDGSVAGNPQDIHQPG 548
 Db 452 XXXXXXXXXXXXXXXXXXXXSRPIFYNYVHPKGAIDVINECYDGSVAGNPQDIHQPG 511
 Qy 549 FAFSRNGPVKRTPTIHLVCRPKRTKASMBFLESEDEGEVQQRYSGGHNLRYFHSDDTC 608
 Db 512 FAFSRNGPVKRTPTIHLVCRPKRTKASMBFLESEDEGEVQQRYSGGHNLRYFHSDDTC 571
 Qy 609 LPLRPQEMEVDSEDEKDPWLREKTTITQIEEFSVDNVEGEKVMKLNHLVHMKGHFIADNQ 668
 Db 572 LPLRPQEMEVDSEDEKDPWLREKTTITQIEEFSVDNVEGEKVMKLNHLVHMKGHFIADNQ 631
 Qy 669 MNHACMLFVENYQCKIIKKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOQKLEKG 728
 Db 632 MNHACMLFVENYQCKIIKKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOQKLEKG 691

QY 729 ESASPANBEITEQNGTANGSEINSEKSKALETDSVSGVSKQKQKL 776
 Db 692 ESASPANBEITEQNGTANGSEINSEKSKALETDSVSGVSKQKQKL 739

RESULT 4
 AAU15978
 ID AAU15978 standard; protein; 388 AA.
 AC AAU15978;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human novel secreted protein, Seq ID 931.
 XX
 KW Human; immunosuppressive; antiarthritic; antiinflammatory; cytostatic;
 KW cardiac; vasotropic; cerebroprotective; neurotropic; neuroprotective;
 KW antibacterial; virucide; fungicide; ophthalmological; vulnery;
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischemia; angioneurosis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;
 KW preservative; antiproliferative.
 OS Homo sapiens.
 XX
 PN WO20015322-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US001341.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 11-JUL-2000; 2000US-0216880P.
 PR 07-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235835P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.

PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488783/53.
N-PSDB; AAS25965.
New nucleic acid molecules encoding 461 human secreted proteins for
diagnosing, preventing, treating or ameliorating medical conditions and
used as food additives or preservatives.
Claim 11; SEQ ID NO 931; 980pp; English.
The invention relates to isolated nucleic acid molecules and their
encoded secreted proteins. The nucleic acids and proteins are used to
prevent, treat or ameliorate a medical condition in e.g. humans, mice,
rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
in diagnosing a pathological condition or susceptibility to a
pathological condition. Antibodies to the proteins can also be used in
alleviating symptoms associated with the disorders and in diagnostic
immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
(ELISA). Disorders which are diagnosed or treated include autoimmune
diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
nervous system disorders e.g. Alzheimer's disease, infections caused by
bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
and many other disorders listed in the specification. The polypeptides
can also be used to aid wound healing and epithelial cell proliferation,
to prevent skin aging due to sunburn, to maintain organs before
transplantation, for supporting cell culture of primary tissues, to
regenerate tissues and in chemotaxis. The polypeptides can also be used
as a food additive or preservative to increase or decrease storage
capabilities, fat content, lipid, protein, carbohydrate, vitamins,
minerals, cofactors and other nutritional components. The present
sequence represents a novel secreted protein of the invention. Note: The
sequence data for this patent did not form part of the printed
Query Match 50.1%; Score 2043; DB 4; Length 388;
Best Local Similarity 99.5%; Pred. No. 4.9e-167;
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 389 GPTLFTLWGTGNDKSTAPAKPLATNSESLSHQENKPGSVKPTQTIAVKESLTTDLQ 448
DB 1 GTTQFTLWGTGNDKSTAPAKPLATNSESLSHQENKPGSVKPTQTIAVKESLTTDLQ 60
QY 449 TRKEKDTFNNRQKRLIFQFLYNNNTROQTEARDLHCPWCTLNCRKLYSLKHLKCH 508

Db 61 TRKEKDTFNNRQKRLIFQFLYNNNTROQTEARDLHCPWCTLNCRKLYSLKHLKCH 120
QY 509 SRFIFNYVHPKCGARIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPTITHLVC 568
Db 121 SRFIFNYVHPKCGARIDVSINEGYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPTITHLVC 180
QY 569 RPKRTKASMEFLESEDEGEVEQOQRTYSSGHNRILYFHSDDTCLPLRQPMEDVSDKDP 628
Db 181 RPKRTKASMEFLESEDEGEVEQOQRTYSSGHNRILYFHSDDTCLPLRQPMEDVSDKDP 240
QY 629 LREKTIQIEEFSDVNEGEKEMKLNHLVHKHGFADNOMHACMLFVENVYQKLIKK 688
Db 241 LREKTIQIEEFSDVNEGEKEMKLNHLVHKHGFADNOMHACMLFVENVYQKLIKK 300
QY 689 LCRNFMHLVSMHDFNLISIMSDKAVTKLRNQKLEKESASPAANEETEEQNGTANG 748
Db 301 LCRNFMHLVSMHDFNLISIMSDKAVTKLRNQKLEKESASPAANEETEEQNGTANG 360
QY 749 FSEINSKEKALETDSVSGVSKQSKOKL 776
Db 361 FSEINSKEKALETDSVSGVSKQSKOKL 388
RESULT 5
ABUS5047
ID ABUS5047 standard; protein; 388 AA.
XX AC ABUS5047;
XX DT 18-MAR-2003 (first entry)
XX DE Human novel polypeptide #134.
KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX OS Homo sapiens.
XX PN US2002132753-A1.
XX PD 19-SEP-2002.
XX PF 17-JAN-2001; 2001US-00764864.
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.
XX PR 14-JUL-2000; 2000US-0218290P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 26-JUL-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225757P.
XX PR 22-AUG-2000; 2000US-0226868P.
XX PR 30-AUG-2000; 2000US-0228924P.
XX PR 01-SEP-2000; 2000US-0229287P.
XX PR 01-SEP-2000; 2000US-0229343P.
XX PR 01-SEP-2000; 2000US-0229344P.
XX PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 17-NOV-2000; 2000US-0244617P.
 PR 01-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2003-147444/14.
 DR N-PSDB; ABX73306.
 XX
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX
 PS Claim 11; SEQ ID NO 931; 402bp; English.
 XX
 CC The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
 CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
 CC ABU55748 represent human novel polypeptides of the invention
 XX
 SQ Sequence 388 AA;
 Query Match 50.1%; Score 2043; DB 6; Length 388;
 Best Local Similarity 99.5%; Pred. No. 4.9e-167;
 Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 389 GPTTQFTLRWGETNDKSTAPIAFLATRNSESILHQENKPSVKPTQTIAVKESLTLDLQ 448
 DB 1 GPTTQFTLRWGETNDKSTAPIAFLATRNSESILHQENKPSVKPTQTIAVKESLTLDLQ 60
 QY 449 TRKEKDTNENRQKLRIFYQFLYNNNTNQQTAEADDLHCPWCTLNCRLKLSLLKHLKCH 508
 DB 61 TRKEKDTNENRQKLRIFYQFLYNNNTNQQTAEADDLHCPWCTLNCRLKLSLLKHLKCH 120

QY 509 SRFIENYVHPKARIDVSIINECYDGSYAGNPQDIHQPGFAFSRNGPVGKTPITHILVC 568
 DB 121 SRFIENYVHPKARIDVSIINECYDGSYAGNPQDIHQPGFAFSRNGPVGKTPITHILVC 180
 QY 569 RPRTKASMSSEFLESDGEVEQORTYSSGHNLRYFHSDTCLPLRQEMEVSEDEKDPDW 628
 DB 181 RPRTKASMSSEFLESDGEVEQORTYSSGHNLRYFHSDTCLPLRQEMEVSEDEKDPDW 240
 QY 629 LREKTIITQIEEFSVDNVEGEKVMKLNHLVVMKHGFIADNMHACMLFVENYGGKIIKKN 688
 DB 241 LREKTIITQIEEFSVDNVEGEKVMKLNHLVVMKHGFIADNMHACMLFVENYGGKIIKKN 300
 QY 689 LCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOKLEKESASPANBEITEEQNGTANG 748
 DB 301 LCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOKLEKESASPANBEITEEQNGTANG 360
 QY 749 FSEINSKEKALETDSVSGVSKQSKQKL 776
 DB 361 FSEINSKEKALETDSVSGVSKQSKQKL 388
 RESULT 6
 AAU15958
 ID AAU15958 standard; protein; 289 AA.
 XX
 AC AAU15958;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human novel secreted protein, Seq ID 911.
 XX
 KW Human; immunosuppressive; antiarthritic; anirrhematic; cytostatic;
 KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
 KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;
 KW preservative; antiproliferative.
 XX
 OS Homo sapiens.
 XX
 PN WO200155322-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US001341.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0188874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 14-JUL-2000; 2000US-0217496P.
 PR 26-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0223268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 16-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0233968P.
PR 14-SEP-2000; 2000US-0233977P.
PR 14-SEP-2000; 2000US-0233988P.
PR 14-SEP-2000; 2000US-0233998P.
PR 14-SEP-2000; 2000US-0234000P.
PR 14-SEP-2000; 2000US-0234001P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.

PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249399P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488783/53.
N-PSDB; AAS25945.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
diagnosing, preventing, treating or ameliorating medical conditions and
used as food additives or preservatives.
XX
XX Claim 11; SEQ ID NO 911; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
encoded secreted proteins. The nucleic acids and proteins are used to
prevent, treat or ameliorate a medical condition in e.g. humans, mice,
rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
in diagnosing a pathological condition or susceptibility to a
pathological condition. Antibodies to the proteins can also be used in
alleviating symptoms associated with the disorders and in diagnostic
immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
(ELISA). Disorders which are diagnosed or treated include autoimmune
diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodysplasia,
nervous system disorders e.g. Alzheimer's disease, infections caused by
bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
and many other disorders listed in the specification. The polypeptides
can also be used to aid wound healing and epithelial cell proliferation,
to prevent skin aging due to sunburn, to maintain organs before
transplantation, for supporting cell culture of primary tissues, to
regenerate tissues and in chemotaxis. The polypeptides can also be used
as a food additive or preservative to increase or decrease storage

CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Query Match 36.7%; Score 1497; DB 4; Length 289;
 Best Local Similarity 99.6%; Pred. No. 3.7e-120;
 Matches 283; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 493 NCRKLYSLKHLKLSCHSRPIFYVYHPKGRIDVINECDGVSAGNPQDIHQPGFAPS 552
 DB 6 DCRKLYSLKHLKLSCHSRPIFYVYHPKGRIDVINECDGVSAGNPQDIHQPGFAPS 65

QY 553 RNPVKTPTIHLVCRPKRTKASMSFLESDGEVEQQRTYSSGHNLVYFSDTCLPLR 612
 DB 66 RNPVKTPTIHLVCRPKRTKASMSFLESDGEVEQQRTYSSGHNLVYFSDTCLPLR 125

QY 613 PQMEVDSDEKDPKWLREKTIITOLEPDSVNEGEKVMKLNHLVWKGFIADNQNHA 672
 DB 126 PQMEVDSDEKDPKWLREKTIITOLEPDSVNEGEKVMKLNHLVWKGFIADNQNHA 185

QY 673 CMLFVNYGQKIIKKNLCRNFMLHLSVMDHFNLSIDKAVTKLREMQQKLEKGSAS 732
 DB 186 CMLFVNYGQKIIKKNLCRNFMLHLSVMDHFNLSIDKAVTKLREMQQKLEKGSAS 245

QY 733 PANBEITEQNTANGFSEINSKEKALETDSVSGVSKQSKKL 776
 DB 246 PANBEITEQNTANGFSEINSKEKALETDSVSGVSKQSKKL 289

RESULT 7
 ABUS5027
 ID ABUS5027 standard; protein; 289 AA.
 XX AC ABUS5027;
 XX DT 18-MAR-2003 (first entry)
 XX DE Human novel polypeptide #114.
 XX KW Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX OS Homo sapiens.
 XX EN US2002132753-A1.
 XX PD 19-SEP-2002.
 XX PF 17-JAN-2001; 2001US-00764864.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225575P.

PR 14-AUG-2000; 2000US-0225758P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2003-147444/14.
 DR N-PSDB; ABX73286.
 XX New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX Claim 11; SEQ ID NO 911; 402pp; English.
 XX The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
 CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and
 CC ABUS5748 represent human novel polypeptides of the invention
 XX Sequence 289 AA;
 SQ

Query Match 36.7%; Score 1497; DB 6; Length 289;
 Best Local Similarity 99.6%; Pred. No. 3.7e-120;
 Matches 283; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 493 NCRKLYSLKHLKLSCHSRPIFYVYHPKGRIDVINECDGVSAGNPQDIHQPGFAPS 552

PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249242P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-488783/53.
 DR N-P5DB; AAS26403.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX
 PS Claim 11; SEQ ID NO 1369; 980pp; English.
 PS
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodysplasia,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,

CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 Query Match 35.6%; Score 1453; DB 4; Length 292;
 Best Local Similarity 96.8%; Pred. No. 2.3e-116;
 Matches 275; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
 QY 233 RKDVSCPIRQVPTGKKQVPLPDLNQTGPNPSPSLAVSSNEFSPNSHMVKSYSLLFRVT 292
 DB 1 KDVSCPIXQVPTGKQXQPLNQLNQTGPNPSPSLAVSSNEFSPNSHMVKSYSLLFRVT 60
 QY 293 RGRREFNGMNGETNENIDVNEELPARRKENREDGKTFVAQMTVFQKNERLQLLDGEY 352
 DB 61 RGRREFNGMNGETNENIDVNEELPARRKENREDGKTFVAQMTVFQKNERLQLLDGEY 120
 QY 353 EVAMQEMEECPISKRAWTETILDGKRLPPFETFSQGTLOFTLRWGTNDKSTAPIAK 412
 DB 121 XVAMQEMEECPISKRAWTETILDGKRLPPFETFSQGTLOFTLRWGTNDKSTAPIAK 180
 QY 413 PLATNSLSLHGENKPGSVKPTQTTIAVKSLETTDLQTRKEKDTNENRQKLRIFYQFLYN 472
 DB 181 PLATNSLSLHGENKPGSVKPTQTTIAVKSLETTDLQTRKEKDTNENRQKLRIFYQFLYN 240
 QY 473 NNTRQOTEARDDLHCPWCTLNCRKLYSLKHLKCHSRFIFNYV 516
 DB 241 NNTRQOTEARDDLHCPWCTLNCRKLYSLKHLKCHSRFIFNXV 284
 RESULT 9
 ABUS5485
 ID ABUS5485 standard; protein; 292 AA.
 XX AC ABUS5485;
 XX DT 18-MAR-2003 (first entry)
 XX DE Human novel polypeptide #572.
 XX KW Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX OS Homo sapiens.
 XX US2002132753-A1.
 XX PD 19-SEP-2002.
 XX PF 17-JAN-2001; 2001US-00764864.
 XX PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.

QY 65 MTDARREQESLKKIKQKLSLTLSSSVRGV-----STPPRHSSGSLTPVTPPTPSS 120
 Db 1 MAPAKKREKDS-----NPDGSAANGILGTHGAPDASNAGSTVPTAGQVKLN 49
 QY 121 SFRSSTP-----TEPTQIYRFLRNLIAPIFLHRTILTYMHSRNTNIKRKTKFYDDM 174
 Db 50 GHQOQELFLQAFKPTQIYLRNRHETNPIFLNRTLSYMKERMSRNNKKRISFQVNSM 109
 QY 175 LSKVEMKGEQESHLSA-HLQFTGFFHKNDKPSSENEQ-----NSVTL 222
 Db 110 LESI-----TOKSEAVSONYLHVYDSLHEKLPARLDNESGEDLLQQLCEAGESVSE 164
 QY 223 VLLVVKCHKRDVSCPTQVPTGKKQVPLIPDLNQTGPNFPLAVSSNEFEP-SNSHM 281
 Db 165 TLYLITRSKKDSTLDQELLKSCQIVYNP---KORVGEHATISIPLOTMRPMGEQHT 221
 QY 282 VKSYSLFRVTRPGRRENGINGETNENIDVNELPARRKRNREDGKTFVAQMTVDFK 341
 Db 222 L--YKLLERIK-----VLSPSTCN--DENAETPP-NKRSRPN-EKMGFSGSLIYK 266
 QY 342 NERLQDLGGEVAMQEMEECPI---SKRATWETILDGKLP---PFETFSQGPPTLOFT 395
 Db 267 SSGP-ITGEVEAMLOPNSISIKSFSPKCTWTMPD-SYPLSLTYDYVYQSPMLKFH 324
 QY 396 LRWTGETNDKSTAPIAKPLATR-----NSESLEHON-----KPGSVKPTQTI 437
 Db 325 LTL---SNEQLPEMISAPELQRYVQHLDVAEMNYYNNNNNNNCSGLKNGSGGNSV 381
 QY 438 AVKESLTTDLQTRKEKDPNENRQKLRIFYOLYNNNTROTEARDLHCPWCITLNCRL 497
 Db 382 C-----KTPP-----EHQIVYFMYSNNTRQOTETQELNCPWGLDCLRL 423
 QY 498 YSLKHLKLSRRIFFNTVYHPKARIDVSNIECYDGSYAGNPQDIHRQPGFAPSRN-GP 556
 Db 424 YALLKHLKLCARFNFTYQAGSGARIDVTINDAYDGSYAGSPYDLAGPSSSPARTCGP 483
 QY 557 VKRPTIHLVCRPKRTKASMEFLESDEGEVQOORTYSSGHNRLYFHSDTCLPLRQEM 616
 Db 484 VRRTSVTSLMVCRRPQKTCLEFLELDEDEISNQRSYITGHNLHYHTTCLPVPHEKL 543
 QY 617 EVDSEKDPWLREKTTQIEEFSDVNKEGVKMLNHLVHMKHGFADNMHACMLF 676
 Db 544 DIDSEGSDDLWLRQKTTQMDIDESDVNKEGKELKMLNHLVHMHGFGVDCQLPIACMP 603
 QY 677 VYENYQKIKKNCNFMHLVSHVDHFNLSIMSDIKAVTKLRMOKLEKGES----- 730
 Db 604 LDKAGTEIVRKNLYRNFILHCSLFDYGLIAETVYKTVQKLOGLLSKYAAGBELMQOR 663
 QY 731 -----ASPANEITE----- 740
 Db 664 BEQLKWLVDGMHKKQBDPKTKSPQKAPPADQASTSSASTSGSGSGSSSQPPKMPA 723
 QY 741 -----EQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776
 Db 724 HLKRGSAASSPGVQSGKTENGTSNS-----SSNSKQVAKKSADQPL 767

RESULT 11

AAU16402
 ID AAU16402 standard; protein; 175 AA.
 XX
 AC AAU16402;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human novel secreted protein, Seq ID 1355.
 XX
 KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
 KW cardiac; vasotropic; cerebroprotective; nootropic; neuroprotective;
 KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
 KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;

KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; skin ageing; food additive;
 XX preservative; antiproliferative.

OS Homo sapiens.

XX WO200155322-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001341.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226686P.

PR 22-AUG-2000; 2000US-0227182P.

KW

KW

XX

OS

XX

XX

XX

PD

XX

PF

XX

XX

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0241826P.
 PR 08-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0244674P.
 PR 08-NOV-2000; 2000US-024475P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246479P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246539P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPT; 2001-488783/53.
 XX N-PSDE; AAS26389.
 DR New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX Claim 11; SEQ ID NO 1355; 980pp; English.
 XX The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Query Match 22.5%; Score 919; DB 4; Length 175;
 Best Local Similarity 94.3%; Pred. No. 1.1e-70;
 Matches 165; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
 Qy 466 FYQFLYNNTRQQTTEARDLLHCPWCTLNCKRLKLLKHLKLSRIFNVVHPKGARID 525
 Db 1 FYQFLYNNTRQQTTEARDLLHCPWCTLNCKRLKLLKHLKLSRIFNVVHPKGARID 60
 Qy 526 VSINECYDGSYAGNPQDIHQPGFAPSRNGPVKRTPTIHLVCRPKRTKASMEFLESD 585
 Db 61 VSINECYDGSYAGNPQDIHQPGFAPSRNGPVKRTPTIHLVCRPKRTKASMEFLESD 120
 Qy 586 GEVEQORTYSSGHNRHLYEHSDTCLPQPOMEVDSDEKDPWLREKTIQTIEEP 640
 Db 121 GEVEQORTYSSGHNRHLYEHSDTCLPQPOMEVDSDEKDPWLREKTIQTIEEP 175

RESULT 12
 ABUS5471
 ID ABUS5471 standard; protein; 175 AA.
 XX AC ABUS5471;
 XX DT 18-MAR-2003 (first entry)
 XX Human novel polypeptide #558.
 XX Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;

KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX Homo sapiens.
 XX US2002132753-A1.
 XX 19-SEP-2002.
 XX 17-JAN-2001; 2001US-00764864.
 XX 31-JAN-2000; 2000US-0179065P.
 XX 04-FEB-2000; 2000US-0180628P.
 XX 28-JUN-2000; 2000US-0214886P.
 XX 07-JUL-2000; 2000US-0216647P.
 XX 07-JUL-2000; 2000US-0216880P.
 XX 11-JUL-2000; 2000US-0217487P.
 XX 11-JUL-2000; 2000US-0217496P.
 XX 14-JUL-2000; 2000US-0218290P.
 XX 26-JUL-2000; 2000US-0220963P.
 XX 26-JUL-2000; 2000US-0220964P.
 XX 14-AUG-2000; 2000US-0224518P.
 XX 14-AUG-2000; 2000US-0224519P.
 XX 14-AUG-2000; 2000US-0225267P.
 XX 14-AUG-2000; 2000US-0225268P.
 XX 14-AUG-2000; 2000US-0225270P.
 XX 14-AUG-2000; 2000US-0225447P.
 XX 14-AUG-2000; 2000US-0225757P.
 XX 22-AUG-2000; 2000US-0225758P.
 XX 30-AUG-2000; 2000US-0226868P.
 XX 01-SEP-2000; 2000US-0228924P.
 XX 01-SEP-2000; 2000US-0229287P.
 XX 01-SEP-2000; 2000US-0229343P.
 XX 01-SEP-2000; 2000US-0229344P.
 XX 05-SEP-2000; 2000US-0229345P.
 XX 05-SEP-2000; 2000US-0229509P.
 XX 08-SEP-2000; 2000US-0229513P.
 XX 21-SEP-2000; 2000US-0231413P.
 XX 21-SEP-2000; 2000US-0234223P.
 XX 21-SEP-2000; 2000US-0234274P.
 XX 25-SEP-2000; 2000US-0234997P.
 XX 27-SEP-2000; 2000US-0235834P.
 XX 29-SEP-2000; 2000US-0236327P.
 XX 29-SEP-2000; 2000US-0236367P.
 XX 29-SEP-2000; 2000US-0236368P.
 XX 29-SEP-2000; 2000US-0236369P.
 XX 29-SEP-2000; 2000US-0236370P.
 XX 02-OCT-2000; 2000US-0236802P.
 XX 02-OCT-2000; 2000US-0237037P.
 XX 02-OCT-2000; 2000US-0237038P.
 XX 02-OCT-2000; 2000US-0237039P.
 XX 02-OCT-2000; 2000US-0237040P.
 XX 13-OCT-2000; 2000US-0239935P.
 XX 20-OCT-2000; 2000US-0240960P.
 XX 20-OCT-2000; 2000US-0241785P.
 XX 20-OCT-2000; 2000US-0241809P.
 XX 01-NOV-2000; 2000US-0244617P.
 XX 17-NOV-2000; 2000US-0249299P.
 XX 08-DEC-2000; 2000US-0251856P.
 XX 08-DEC-2000; 2000US-0251868P.
 XX 08-DEC-2000; 2000US-0251869P.
 XX (ROSE/) ROSEN C A.
 XX (RUBE/) RUBEN S M.
 XX (BARA/) BARASH S C.
 XX Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2003-147444/14.
 XX N-PSDB; ABX73730.
 XX New polypeptides and nucleic acids, useful in gene therapy for treating,
 XX inhibiting or preventing e.g. neural, immune system, muscular,

PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX Claim 11; SEQ ID NO 1355; 402pp; English.
 XX The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
 CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and
 CC ABUS5748 represent human novel polypeptides of the invention
 XX Sequence 175 AA;
 XX
 Query Match 22.5%; Score 919; DB 6; Length 175;
 Best Local Similarity 94.3%; Pred. No. 1.1e-70;
 Matches 165; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
 QY 466 FYQFLYNNNTROQTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIYNYVHPKGARID 525
 DB 1 FYQFLYNNNTROQTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIYNYVHPKGARID 60
 QY 526 VSINECYDGSYAGNPQDIHQPGFAFSRNGPVKRTPIITHILVCRPKTKASMSFELED 585
 DB 61 VSINECYDGSYAGNPQDIHQPGFAFSRNGPVKRTPIITHILVCRPKTKASMSFELED 120
 QY 586 GEVEQRTYSSGHNRKLYPHSDTCLPLRPQMEVDSDEKDPFWLREKTIQIEEF 640
 DB 121 GEVEQRTYSSGHNRKLYPHSDTCLPLRPQMEVDSDEKDPFWLREKTIQIEEF 175
 RESULT 13
 ADB64250
 ID ADB64250 standard; protein; 243 AA.
 AC ADB64250;
 XX
 DT 04-DEC-2003 (first entry)
 XX Human protein encoded by clone DFNES20076340.
 DE
 XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
 KW cell regeneration; membrane protein; signal transduction-related protein;
 KW transcription-related protein; osteoporosis; neurological disease;
 KW cancer; tumour.
 XX Homo sapiens.
 XX EP1308459-A2.
 XX 07-MAY-2003.
 XX 28-MAR-2002; 2002EP-00007401.
 XX 05-NOV-2001; 2001JP-00379298.
 XX 25-JAN-2002; 2002US-00350978.
 XX (HELI-) HELIX RES INST.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX Isogai T, Sugiyama T, Otsuka T, Wakamatsu A, Sato H, Ishii S;
 XX Yamanoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX

XX PD 02-AUG-2001.
 XX PF 17-JAN-2001; 2001WO-US001341.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226273P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 23-AUG-2000; 2000US-0227182P.
 PR 30-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.

PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX (HUYA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 XX
 XX

DR WPI: 2001-488783/53.
DR N-PSDB; AAS26368.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX
XX Claim 11; SEQ ID NO 1334; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 16.0%; Score 652; DB 4; Length 278;
Best Local Similarity 99.2%; Pred No. 2.5e-47;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTGIAAAGFFSNTCRFGGGLHFTLADLIEHDNHDIDTPRVLEKQLQPTYVALSY 60
Db |||||
36 MTGIAAAGFFSNTCRFGGGLHFTLADLIEHDNHDIDTPRVLEKQLQPTYVALSY 95
QY 61 INRFMTDAARREOESLKKIOPKLSLTLSVVSGNVSTPPRHSSGSLTPVTPPTPSS 120
Db |||||
96 INRFMTDAARREOESLKKIOPKLSLTLSVVSGNVSTPPRHSSGSLTPVTPPTPSS 155
QY 121 SFRSSTPT 128
Db |||||
156 SFRSSTPT 163

Search completed: August 25, 2004, 17:54:35
Job time : 137 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 18:06:05 ; Search time 134 Seconds
(without alignments)
1827.180 Million cell updates/sec

Title: US-09-874-162A-8
Perfect score: 776
Sequence: 1 MTGIAAASFNTCRFGCG.....KALETDSVSGVSKQKQKL 776

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 9 | 1.2 | 1170 | 5 Q95T15 | Q95T15 drosophila |
| 2 | 9 | 1.2 | 1990 | 5 Q9U8Q0 | Q9U8Q0 drosophila |
| 3 | 9 | 1.2 | 1995 | 5 Q9W244 | Q9W244 drosophila |
| 4 | 8 | 1.0 | 75 | 10 Q9STQ5 | Q9STQ5 arabidopsis |
| 5 | 8 | 1.0 | 93 | 5 Q61719 | Q61719 anopheles g |
| 6 | 8 | 1.0 | 150 | 10 Q3FP19 | Q3FP19 oryza sativ |
| 7 | 8 | 1.0 | 150 | 10 Q8H2H9 | Q8H2H9 oryza sativ |
| 8 | 8 | 1.0 | 156 | 10 Q9LEY9 | Q9LEY9 arabidopsis |
| 9 | 8 | 1.0 | 214 | 10 P93540 | P93540 silene prat |
| 10 | 8 | 1.0 | 214 | 10 Q9ST39 | Q9ST39 silene lati |
| 11 | 8 | 1.0 | 217 | 10 O04301 | O04301 silene lati |
| 12 | 8 | 1.0 | 217 | 10 P93689 | P93689 silene lati |
| 13 | 8 | 1.0 | 247 | 12 Q8B9D8 | Q8B9D8 rachiplusia |
| 14 | 8 | 1.0 | 281 | 4 Q9EQW4 | Q9EQW4 homo sapien |
| 15 | 8 | 1.0 | 281 | 16 Q8XBX6 | Q8XBX6 escherichia |
| 16 | 8 | 1.0 | 297 | 16 Q8NPQ7 | Q8NPQ7 corynebacte |

| | | | | | |
|----|---|-----|------|-----------|---------------------|
| 17 | 8 | 1.0 | 305 | 16 Q81E61 | Q81E61 bacillus ce |
| 18 | 8 | 1.0 | 307 | 17 Q8U0U5 | Q8U0U5 pyrococcus |
| 19 | 8 | 1.0 | 316 | 10 Q9LS48 | Q9LS48 arabidopsis |
| 20 | 8 | 1.0 | 340 | 4 Q9BRF4 | Q9BRF4 homo sapien |
| 21 | 8 | 1.0 | 342 | 4 Q9Y6J7 | Q9Y6J7 homo sapien |
| 22 | 8 | 1.0 | 385 | 4 Q9BQW7 | Q9BQW7 homo sapien |
| 23 | 8 | 1.0 | 442 | 17 Q9UYR0 | Q9UYR0 pyrococcus |
| 24 | 8 | 1.0 | 447 | 16 Q895X3 | Q895X3 clostridium |
| 25 | 8 | 1.0 | 476 | 13 Q13256 | Q13256 gallus gall |
| 26 | 8 | 1.0 | 478 | 5 Q95YB9 | Q95YB9 caenorhabdi |
| 27 | 8 | 1.0 | 503 | 13 Q9CLB3 | Q9CLB3 pasteurella |
| 28 | 8 | 1.0 | 506 | 13 Q9DEF6 | Q9DEF6 gallus gall |
| 29 | 8 | 1.0 | 508 | 13 Q9DEF5 | Q9DEF5 gallus gall |
| 30 | 8 | 1.0 | 514 | 13 Q13255 | Q13255 gallus gall |
| 31 | 8 | 1.0 | 520 | 4 Q81XL3 | Q81XL3 homo sapien |
| 32 | 8 | 1.0 | 525 | 10 Q8H133 | Q8H133 arabidopsis |
| 33 | 8 | 1.0 | 526 | 10 Q98ZQ3 | Q98ZQ3 arabidopsis |
| 34 | 8 | 1.0 | 668 | 4 Q9BR66 | Q9BR66 homo sapien |
| 35 | 8 | 1.0 | 697 | 16 Q7VQM0 | Q7VQM0 candidatus |
| 36 | 8 | 1.0 | 881 | 5 Q81SG6 | Q81SG6 stylonychia |
| 37 | 8 | 1.0 | 3525 | 5 Q81SW5 | Q81SW5 plasmodium |
| 38 | 7 | 0.9 | 51 | 16 Q8FPP5 | Q8FPP5 corynebacte |
| 39 | 7 | 0.9 | 51 | 16 Q7ULM1 | Q7ULM1 rhodospirell |
| 40 | 7 | 0.9 | 52 | 16 P74799 | P74799 synechocyst |
| 41 | 7 | 0.9 | 56 | 12 Q7T7N2 | Q7T7N2 human picob |
| 42 | 7 | 0.9 | 70 | 16 Q8DD29 | Q8DD29 vibrio vuln |
| 43 | 7 | 0.9 | 76 | 16 Q81HA1 | Q81HA1 bacillus ce |
| 44 | 7 | 0.9 | 84 | 16 Q87HG7 | Q87HG7 vibrio para |
| 45 | 7 | 0.9 | 88 | 16 Q8R7X6 | Q8R7X6 thermoanaer |

ALIGNMENTS

RESULT 1

| | | | | | |
|--------|---|--------------------------------------|--------------|------|----------|
| Q95T15 | ID | Q95T15 | PRELIMINARY; | PRT; | 1170 AA. |
| AC | Q95T15 | 01-DEC-2001 (TREMBLrel. 19, Created) | | | |
| DT | 01-DEC-2001 (TREMBLrel. 19, Last sequence update) | | | | |
| DB | 01-OCT-2002 (TREMBLrel. 22, Last annotation update) | | | | |
| DS | SD01586P. | | | | |
| GN | PX OR CG4444. | | | | |
| OS | Drosophila melanogaster (Fruit fly). | | | | |
| OC | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | | | |
| OC | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | | | |
| OX | NCBI_TaxID=7227; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RA | Scapleton M., Brokstein P., Hong L., Agbayani A., Carlson J., | | | | |
| RA | Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R., | | | | |
| RA | Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J., | | | | |
| RA | Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K., | | | | |
| RA | Yu C., Lewis S.E., Rubin G.M., Celnik S. | | | | |
| EL | Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases. | | | | |
| DR | EMBL; AY058753; AAL13982.1; - | | | | |
| DR | FlyBase; FBgn0003175; px. | | | | |
| SQ | SEQUENCE 1170 AA; 124080 MW; 37F9331CF6891DD3 CRC64; | | | | |

Query Match 1.2%; Score 9; DB 5; Length 1170;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | |
|----|-----|---------------|
| Qy | 113 | TPFITPSSS 121 |
| Db | 907 | TPFITPSSS 915 |

RESULT 2

| | | | | | |
|--------|--------|--------|--------------|------|----------|
| Q9U8Q0 | ID | Q9U8Q0 | PRELIMINARY; | PRT; | 1990 AA. |
| AC | Q9U8Q0 | | | | |

DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
 DE PLEXUS.
 GN PX OR CG4444.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=20025921; PubMed=10556047;
 RA Matkatsuo H., Tadokoro R., Gano S., Hayashi S.;
 RT "Regression of the wing vein development in Drosophila by the nuclear
 RT matrix protein Plexus.";
 RL Development 126:5207-5216(1999).
 DR EMBL; AB032181; BAA86717.1; -;
 DR FlyBase; FBgn0003175; px.
 SQ SEQUENCE 1990 AA; 213113 MW; 5707C4F2986DA9E2 CRC64;

 Query Match 1.2%; Score 9; DB 5; Length 1990;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 113 TPPITPSSS 121
 Db 1727 TPPITPSSS 1735

 RESULT 3
 Q9W244 PRELIMINARY; PRT; 1995 AA.
 AC Q9W244;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
 DT 01-WAR-2003 (TREMELrel. 23, Last annotation update)
 DE CG4444 protein.
 GN PX OR CG4444.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Beran B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.N.,
 RA Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J.R., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Douc L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jallali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Patel S., Pfeiffer B.,
 RA Pacleb J., Paragas V., Park S., Patel S., Richards S., Scheeler F.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003457; AAF46855.2; -;
 DR FlyBase; FBgn0003175; px.
 SQ SEQUENCE 1995 AA; 213862 MW; E7A563275E1D8C7 CRC64;

 Query Match 1.2%; Score 9; DB 5; Length 1995;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 113 TPPITPSSS 121
 Db 1732 TPPITPSSS 1740

 RESULT 4
 Q9STQ5 PRELIMINARY; PRT; 75 AA.
 AC Q9STQ5;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-WAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN T25K17.100 OR AT4G26290.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Koetter P., Hempel S., Entian K.-D., Bancroft I.,
 RA Mewes H.W., Mayer K.F.X., Schueller C.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBSJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBSJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AL049171; CAB38958.1; -
 DR EMBL; AL161564; CAB79484.1; -
 DR PIR; T06013; T06013.
 KW Hypothetical protein.
 SQ SEQUENCE 75 AA; 8649 MW; 5842E9469EBD2BD1 CRC64;
 Query Match 1.0%; Score 8; DB 10; Length 75;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 115 PITPSSSF 122
 DB 48 PITPSSSF 55
 RESULT 5
 O61719 PRELIMINARY; PRT; 93 AA.
 AC O61719;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Anopheles gambiae (African malaria mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelinae.
 OX NCBI_TaxID=7165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=G3;
 RA Romans P., Law A.K.-T., Binns K., Han Y.-S., Cornel A.J.,
 RA Vasiliuskas D., Paskewitz S.M.;
 RT "Molecular genetics of catecholamine metabolism in the malaria vector
 RT Anopheles gambiae: analysis of the dopa decarboxylase gene."
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AF063021; AAC16248.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 93 AA; 10569 MW; 8E1BBA930E41AD84 CRC64;
 Query Match 1.0%; Score 8; DB 5; Length 93;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 84 LSLTSSSS 91
 DB 46 LSLTSSSS 53
 RESULT 6
 Q9FP19 PRELIMINARY; PRT; 150 AA.
 ID Q9FP19
 AC Q9FP19;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE P0038C05.4 protein.
 GN P0038C05.4.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
 RT clone:P0038C05."
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AP003044; BAB19331.1; -
 DR HSSP; P55769; 127K.
 DR Gramene; Q9FP19; -
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR004038; Ribosomal_L7A.
 DR InterPro; IPR004037; Ribosomal_L7Ae.
 DR Pfam; PF01248; Ribosomal_L7Ae; 1.
 DR PRINTS; PRO0881; L7ARS6FAMILY.
 SQ SEQUENCE 150 AA; 16354 MW; 6F626F759AE30108 CRC64;
 Query Match 1.0%; Score 8; DB 10; Length 150;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 408 APIAKPLA 415
 DB 17 APIAKPLA 24
 RESULT 7
 Q8H2H9 PRELIMINARY; PRT; 150 AA.
 AC Q8H2H9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE P0676F10.14 protein.
 GN P0676F10.14.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
 RT clone:P0676F10."
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AP005813; BAC20622.1; -
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR004038; Ribosomal_L7A.
 DR InterPro; IPR004037; Ribosomal_L7Ae.
 DR Pfam; PF01248; Ribosomal_L7Ae; 1.
 DR PRINTS; PRO0881; L7ARS6FAMILY.
 SQ SEQUENCE 150 AA; 16354 MW; 6F626F759AE30108 CRC64;
 Query Match 1.0%; Score 8; DB 10; Length 150;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 408 APIAKPLA 415

```

DB      17 APIAKPLA 24
|||||
RESULT 8
Q9LEY9 PRELIMINARY; PRT; 156 AA.
AC Q9LEY9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Nhp2-like protein (AT5G08180/T22D6_120).
GN T22D6.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buyschaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RL "Arabidopsis cDNA clones.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Chung M.K., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yu G., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
RL "Arabidopsis cDNA clones.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL357612; CAB93719.1; -
DR EMBL; AY050477; AAK91490.1; -
DR EMBL; AF378885; AAK55688.1; -
DR FIR; T50503; T50503.
DR HSP; P55769; 1E7K.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; P:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR004038; Ribosomal_L7A.
DR InterPro; IPR004037; Ribosomal_L7Ae.
DR Pfam; PF01248; Ribosomal_L7Ae; 1.
DR PRINTS; PR00891; L7ARSGFAMILY.
SQ SEQUENCE 156 AA; 16948 MW; CC8B7C196E92442A CRC64;

Query Match 1.0%; Score 8; DB 10; Length 156;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 APIAKPLA 415
DB 23 APIAKPLA 30
|||||
RESULT 9
P93540 PRELIMINARY; PRT; 214 AA.
AC P93540;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MROS3 protein precursor.
GN MROS3.
OS Silene pratensis (White campion) (Lychnis alba).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=52853;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Anther;
RX MEDLINE=97048713; PubMed=8893544;
RA Matsunaga S., Kawano S., Takano H., Uchida H., Sakai A., Kuroiwa T.;
RL "Isolation and developmental expression of male reproductive organ-
specific genes in a dioecious campion, Melandrium album (Silene
latifolia).";
RL Plant J. 10:679-689(1996).
DR EMBL; D82027; BAA11513.1; -
DR InterPro; IPR007493; DUF538.
DR Pfam; PF04398; DUF538; 1.
RW Signal.
FT SIGNAL.
FT SIGNAL.
SQ SEQUENCE 214 AA; 23461 MW; 1931BC0C91935AFB CRC64;

Query Match 1.0%; Score 8; DB 10; Length 214;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSILTSSS 91
DB 16 LSILTSSS 23
|||||
RESULT 10
Q9ST39 PRELIMINARY; PRT; 214 AA.
AC Q9ST39;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MROS3A.
GN MROS3A.
OS Silene latifolia (White campion).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=37657;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX Matsunaga S., Kawano S., Kuroiwa T.;
RL "Chromosomal localization of MROS3 genomic clones in a dioecious
campion, Silene latifolia.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=97048713; PubMed=8893544;
RA Matsunaga S., Kawano S., Takano H., Uchida H., Sakai A., Kuroiwa T.;
RL "Isolation and developmental expression of male reproductive organ-
specific genes in a dioecious campion, Melandrium album (Silene
latifolia).";
RL Plant J. 10:679-689(1996).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=SATI; TISSUE=Leaf;
RX MEDLINE=20040083; PubMed=10571898;
RA Matsunaga S., Schutze K., Donnison I.S., Grant S.R., Kuroiwa T.,
RA Kawano S.;
RT "Single pollen typing combined with laser-mediated manipulation.";
RL Plant J. 20:371-378(1999).
DR EMBL; AB013611; BAA84010.1; -.
DR EMBL; AB029398; BAA90868.1; -.
DR InterPro; IPR007493; DUF538.
DR Pfam; PF04398; DUF538; 1.
SQ SEQUENCE 214 AA; 23461 MW; 492EAF0480CD7C20 CRC64;

Query Match 1.0%; Score 8; DB 10; Length 214;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSLTSSS 91
Db 16 LSLTSSS 23

RESULT 11
O04301 PRELIMINARY; PRT; 217 AA.
ID O04301;
AC O04301;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Men-9a protein.
GN MEN-9A.
OS Silene latifolia (white campion).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=37657;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=flower;
RA Robertson S.E., Scutt C.P., Li Y., Willis M.E., Gilmartin P.M.;
RT "Expression dynamics of the men-9 genes delimit the third whorl in
RT dioecious Silene latifolia flowers.";
RL Plant J. 0:0-0(0).
DR EMBL; Y12324; CAA72993.1; -.
DR InterPro; IPR007493; DUF538.
DR Pfam; PF04398; DUF538; 1.
SQ SEQUENCE 217 AA; 23946 MW; 2E3640753B266F0E CRC64;

Query Match 1.0%; Score 8; DB 10; Length 217;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSLTSSS 91
Db 17 LSLTSSS 24

RESULT 12
P93689 PRELIMINARY; PRT; 217 AA.
ID P93689;
AC P93689;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CCL54 protein.
GN CCL54.
OS Silene latifolia (white campion).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=37657;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Anther;

RX MEDLINE=98042964; PubMed=9375394;
RA Barbacar N., Himmelsdaels S., Farbos I., Moneger F., Lardon A.,
RA Delichere C., Mouras A., Negrutiu I.;
RT "Isolation of early genes expressed in reproductive organs of the
RT dioecious white campion (Silene latifolia) by subtraction cloning
RT using an asexual mutant";
RL Plant J. 12:805-817(1997).
DR EMBL; X94358; CAA64143.1; -.
DR InterPro; IPR007493; DUF538.
DR Pfam; PF04398; DUF538; 1.
SQ SEQUENCE 217 AA; 23968 MW; 154F8B059BB80D84 CRC64;

Query Match 1.0%; Score 8; DB 10; Length 217;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSLTSSS 91
Db 17 LSLTSSS 24

RESULT 13
Q8B9D8 PRELIMINARY; PRT; 247 AA.
ID Q8B9D8;
AC Q8B9D8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Rachiplusia ou multiple nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=80366;
RN [1]
RP SEQUENCE FROM N.A.
RA Bonning B.C., Harrison R.L.;
RT "The Rachiplusia ou multiple nucleopolyhedrovirus genome sequence.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
EMBL; AY145471; AAN28079.1; -.
KW Hypothetical protein.
SQ SEQUENCE 247 AA; 28705 MW; FD6943F62F44D7A9 CRC64;

Query Match 1.0%; Score 8; DB 12; Length 247;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 YVALSYIN 62
Db 23 YVALSYIN 30

RESULT 14
Q9BQW4 PRELIMINARY; PRT; 281 AA.
ID Q9BQW4;
AC Q9BQW4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DJ81LH13.1.4 (Phospholipase C, beta 4 (1-phosphatidylinositol-4,5-
DE bisphosphate phosphodiesterase beta 4) (Isoform 4)) (Fragment).
GN PLCB4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Blakey S.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AL023805; CAC34529.1; -.
DR HSSP; P10688; 1DJX.
DR GO; GO:0004435; P:phosphoinositide phospholipase C activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.

```

```

DR GO: GO:0006629; P:lipid metabolism; IEA.
DR InterPro: IPR001192; PI_PLC.
DR InterPro: IPR000909; PI_PLC_Xdom.
DR Pfam: PF00388; PI_PLC-X_1.
DR PRINTS: PR00390; BPHPLIPASEC.
DR SMART: SM00148; PLCXC; 1.
DR PROSITE: PS00007; PIPLC_X_DOMAIN; 1.
FT NON TER      281
SQ SEQUENCE      281 AA; 32062 MW; 66549A020E7D7F6F CRC64;

Query Match      1.0%; Score 8; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 28;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      728 GESASpan 735
DB      216 GESASpan 223
|||||

RESULT 15
Q8XB6
ID Q8XB6 PRELIMINARY; PRT; 281 AA.
AC Q8XB6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative transposase (St46 protein).
GN ST46 OR 24330 OR ECS3859.
OS Escherichia coli, and
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=83/39;
RA Tauschek M., Strugnell R.A., Robins-Browne R.M.;
RT "Characterization of the LEE pathogenicity islands of rabbit
enteropathogenic Escherichia coli.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RP SPECIES=E.coli; STRAIN=413/89-1;
RA Benkel P., Chakraborty T.;
RT "Genetic organisation and sequence of the LEE II locus in Shiga toxin-
producing Escherichia coli.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RP SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Weich R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[4]
SEQUENCE FROM N.A.
RP SPECIES=E.coli O157:H7; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sakakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AF453441; AAL57563.1; -
DR EMBL; AJ277443; CAC81884.1; -
DR EMBL; AE005528; AAG58114.1; -

```

```

DR EMBL; AP002563; BAB37282.1; ALT_INIT.
DR PIR; C91111; C91111.
DR PIR; P85956; P85956.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006310; P:DNA recombination; IEA.
DR InterPro: IPR001584; Rve.
DR Pfam; PF00665; rve; 1.
KW Complete proteome.
SQ SEQUENCE      281 AA; 32894 MW; 1A64154BE8067514 CRC64;

Query Match      1.0%; Score 8; DB 16; Length 281;
Best Local Similarity 100.0%; Pred. No. 28;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      499 SLKHLKL 506
DB      223 SLKHLKL 230
|||||

```

Search completed: August 25, 2004, 18:14:44
Job time : 141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 18:07:10 ; Search time 46 Seconds
(without alignments)
1622.709 Million cell updates/sec

Title: US-09-874-162A-8
Perfect score: 776
Sequence: 1 MTGIAAASFNTCRFGGCG.....KALETDSVSGVSKQKQL 776

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 8 | 1.0 | 75 | T06013 | hypothetical prote |
| 2 | 8 | 1.0 | 156 | T50503 | nbp2-like protein |
| 3 | 8 | 1.0 | 247 | E72865 | AcOrf-124 protein |
| 4 | 8 | 1.0 | 262 | C91111 | hypothetical prote |
| 5 | 8 | 1.0 | 281 | F85956 | probable transposa |
| 6 | 8 | 1.0 | 442 | C75057 | hypothetical prote |
| 7 | 8 | 1.0 | 526 | T08541 | hypothetical prote |
| 8 | 8 | 1.0 | 641 | JC7331 | gamma-glutamyltran |
| 9 | 8 | 1.0 | 669 | T08827 | hypothetical prote |
| 10 | 7 | 0.9 | 46 | 1 VTVA42 | viscotoxin A2 - Eu |
| 11 | 7 | 0.9 | 46 | 1 VFPD1T | phoratoxin - Calif |
| 12 | 7 | 0.9 | 52 | S76642 | hypothetical prote |
| 13 | 7 | 0.9 | 105 | A81046 | SugE protein [impo |
| 14 | 7 | 0.9 | 109 | F97132 | hypothetical prote |
| 15 | 7 | 0.9 | 116 | H90626 | NADH dehydrogenase |
| 16 | 7 | 0.9 | 129 | T00282 | hypothetical prote |
| 17 | 7 | 0.9 | 142 | E84040 | hypothetical prote |
| 18 | 7 | 0.9 | 144 | T50649 | elicitor-responsiv |
| 19 | 7 | 0.9 | 152 | 1 Q0RCE1 | regulatory protein |
| 20 | 7 | 0.9 | 152 | 2 F81214 | transcription regu |
| 21 | 7 | 0.9 | 152 | 2 F80606 | regulator for asna |
| 22 | 7 | 0.9 | 152 | 2 AC0953 | regulatory protein |
| 23 | 7 | 0.9 | 153 | 1 Q4ADB2 | early E4 17K prote |
| 24 | 7 | 0.9 | 154 | 2 T50100 | Nucleolar protein, |
| 25 | 7 | 0.9 | 154 | 2 T43644 | nbp2 homolog - fis |
| 26 | 7 | 0.9 | 155 | 2 B70030 | conserved hypothet |
| 27 | 7 | 0.9 | 157 | 2 S75905 | hypothetical prote |
| 28 | 7 | 0.9 | 164 | 2 T01726 | hypothetical prote |
| 29 | 7 | 0.9 | 190 | 2 AB2305 | 6,7-dimethyl-8-rib |

| | | | | | |
|----|---|-----|-----|----------|--------------------|
| 30 | 7 | 0.9 | 198 | 2 B81389 | hypothetical prote |
| 31 | 7 | 0.9 | 201 | 2 T46404 | hypothetical prote |
| 32 | 7 | 0.9 | 222 | 2 C75539 | conserved hypothet |
| 33 | 7 | 0.9 | 223 | 2 T23356 | hypothetical prote |
| 34 | 7 | 0.9 | 224 | 2 D96643 | hypothetical prote |
| 35 | 7 | 0.9 | 227 | 2 AH2123 | hypothetical prote |
| 36 | 7 | 0.9 | 230 | 2 B86824 | two-component syst |
| 37 | 7 | 0.9 | 234 | 2 S22134 | capsid protein - r |
| 38 | 7 | 0.9 | 237 | 2 S75120 | transposase sll198 |
| 39 | 7 | 0.9 | 239 | 2 T45806 | hypothetical prote |
| 40 | 7 | 0.9 | 247 | 2 F84420 | hypothetical prote |
| 41 | 7 | 0.9 | 251 | 2 E96637 | hypothetical prote |
| 42 | 7 | 0.9 | 254 | 2 T15952 | hypothetical prote |
| 43 | 7 | 0.9 | 260 | 2 T36955 | hypothetical prote |
| 44 | 7 | 0.9 | 269 | 2 T35647 | phosphomethylpyrim |
| 45 | 7 | 0.9 | 269 | 2 AG0718 | probable toxin-lik |

ALIGNMENTS

RESULT 1
T06013
hypothetical protein T25K17.100 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06013
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15382
A:Accession: T06013
A:Molecule type: DNA
A:Residues: 1-75 <BEV>
A:Cross-references: EMBL:AL049171; GSPDB:GN00062; ATSP:T25K17.100
A:Experimental source: cultivar Columbia; BAC clone T25K17
C:Genetics:
A:Gene: ATSP:T25K17.100
A:Map position: 4

Query Match 1.0%; Score 8; DB 2; Length 75;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 115 PTPSSSF 122
Db 48 PTPSSSF 55

RESULT 2
T50503
nbp2-like protein - Arabidopsis thaliana
N:Alternate names: protein T22D6.120
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50503
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25101
A:Accession: T50503
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <BEV>
A:Cross-references: EMBL:AL357612
A:Experimental source: cultivar Columbia; BAC clone T22D6
C:Genetics:
A:Map position: 5
A:Introns: 45/1; 68/2; 103/3
A:Note: T22D6.120

Query Match 1.0%; Score 8; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 APIAKPLA 415
 |||||
 Db 23 APIAKPLA 30

RESULT 3

E72865
 AcOrf-124 protein - Autographa californica nuclear polyhedrosis virus
 C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV
 A:Note: dsDNA virus
 C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 12-Nov-1999
 C:Accession: E72865
 R:Ayres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.
 Virology 202, 566-605, 1994
 A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.
 A:Reference number: A72850; MUID:94303173; PMID:8030224
 A:Accession: E72865
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-247 <AYR>
 A:Cross-references: GB:L22858; NID:gs10708; PIDN:AAA66754.1; PID:gs559193
 C:Genetics:
 A:Gene: ACOrf-124

Query Match 1.0%; Score 8; DB 2; Length 247;
 Best Local Similarity 100.0%; Pred. No. 9.4; 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0

QY 55 YVALSYIN 62
 |||||
 Db 23 YVALSYIN 30

RESULT 4

C9111
 Hypothetical protein Ecs3859 [imported] - Escherichia coli (strain O157:H7, substrain R1
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: C9111
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A:Reference number: A39629; MUID:21156231; PMID:11258796
 A:Accession: C9111
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-262 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BAE37282.1; PID:gl3363331; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: Ecs3859

Query Match 1.0%; Score 8; DB 2; Length 262;
 Best Local Similarity 100.0%; Pred. No. 9.9; 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0

QY 499 SLLKHLKL 506
 |||||
 Db 204 SLLKHLKL 211

RESULT 5

F85956
 Probable transposase Z4330 [imported] - Escherichia coli (strain O157:H7, substrain ED19
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: F85956
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: F85956
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-281 <STO>
 A:Cross-references: GB:AE005174; NID:gl2517538; PIDN:AAGS8114.1; GSPDB:GN00145; UWGP:Z433;
 A:Experimental source: strain O157:H7, substrain ED1933
 C:Genetics:
 A:Gene: Z4330

Query Match 1.0%; Score 8; DB 2; Length 281;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 SLLKHLKL 506
 |||||
 Db 223 SLLKHLKL 230

RESULT 6

C75057
 Hypothetical protein PAB1420 - Pyrococcus abyssi (strain Orsay)
 C:Species: Pyrococcus abyssi
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: C75057

R:anonymous, Genoscope
 submitted to the EMBL Data Library, July 1999
 A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
 A:Reference number: A75001
 A:Accession: C75057
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-442 <KAW>
 A:Cross-references: GB:AJ248287; GB:AL096836; NID:gs458657; PIDN:CAB50352.1; PID:el516255
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: PAB1420

Query Match 1.0%; Score 8; DB 2; Length 442;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 LLDGEYEV 354
 |||||
 Db 135 LLDGEYEV 142

RESULT 7

T08541
 Hypothetical protein F27B13.40 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Apr-2003
 C:Accession: T08541
 R:Bevan, M.; Zimmermann, W.; Gruenisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Maye
 submitted to the Protein Sequence Database, May 1999
 A:Reference number: Z16442
 A:Accession: T08541
 A:Molecule type: DNA
 A:Residues: 1-526 <BEV>
 A:Cross-references: EMBL:AL050352; GSPDB:GN00062; ATSP:F27B13.40
 A:Experimental source: cultivar Columbia; BAC clone F27B13
 C:Genetics:
 A:Gene: ATSP:F27B13.40
 A:Map position: 4
 A:Introns: 421/3
 C:Superfamily: patatin

Query Match 1.0%; Score 8; DB 2; Length 526;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 TPSSSFRS 124
 |||||

Db 109 TPSSFRS 116

RESULT 8

JC7331

gamma-glutamyltransferase (EC 2.3.2.2) homolog - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 15-Sep-2000

C:Accession: JC7331

R:Yamaguchi, T.; Takai, N.; Araki, K.; Ishii, K.; Nagano, T.; Ichikawa, T.; Kumanishi, T.

J. Biochem. 128, 101-106, 2000

A:Title: Molecular characterization of a novel gamma-glutamyl transpeptidase homologue f

A:Reference number: JC7331

A:Accession: JC7331

A:Molecule type: mRNA

A:Residues: 1-641 <YAM>

A:Cross-references: GB:AF244973

A:Experimental source: brain

C:Comment: This enzyme is a membrane-bound glycoenzyme that is involved in the metabolis

glutathione.

C:Genetics:

A:Gene: gtpb

C:Keywords: aminoacyltransferase; brain; glutathione; glycolysis; membrane bound

Query Match 1.0%; Score 8; DB 2; Length 641;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 DMLSKVE 179

|||||

Db 415 DMLSKVE 422

RESULT 9

T08827

hypothetical protein cotel - human

C:Species: Homo sapiens (man)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000

C:Accession: T08827

R:Winfield, S.L.; Tayebi, N.; Martin, B.M.; Ginns, B.I.; Sidransky, B.

Genome Res. 7, 1020-1026, 1997

A:Title: Identification of three additional genes contiguous to the glucocerebrosidase 1

A:Reference number: Z16482; MUID:97474796; PMID:9331372

A:Accession: T08827

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-669 <WIN>

A:Cross-references: EMBL:AF023268; NID:G2564910; PIDN:AAC51822.1; PID:G2564916

C:Genetics:

A:Gene: cotel

A:Map position: 1

A:Introns: 75/3; 94/3; 131/3; 171/3; 207/3; 266/2; 299/3; 323/1; 505/1; 528/1; 612/3

Query Match 1.0%; Score 8; DB 2; Length 669;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 SSGSLTPP 111

|||||

Db 467 SSGSLTPP 474

RESULT 10

VTVAA2

viscotoxin A2 - European mistletoe

C:Species: Viscum album (European mistletoe)

C>Date: 13-Jul-1981 #sequence_revision 08-Oct-1981 #text_change 04-Oct-1996

C:Accession: A90005; A01799

R:Olson, T.; Samuelsson, G.

Acta Chem. Scand. 26, 585-595, 1972

A:Title: The amino acid sequence of viscotoxin A2 from the European mistletoe (Viscum al

A:Reference number: A90005; MUID:72211843; PMID:5035954

A:Accession: A90005

A:Molecule type: protein

A:Residues: 1-46 <OLS>

R:Olson, T.; Samuelsson, G.

Acta Pharm. Suec. 11, 381-386, 1974

A:Title: The disulphide bonds of viscotoxin A2 from the European mistletoe (Viscum album

A:Reference number: A90013; MUID:75015879; PMID:4607177

A:Contents: annotation; disulfide bonds

C:Superfamily: viscotoxin

C:Keywords: toxin

F:3-40 4-32/Disulfide bonds: #status predicted

F:16-26/Disulfide bonds: #status experimental

Query Match 0.9%; Score 7; DB 1; Length 46;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 NTCRFGG 18

|||||

Db 14 NTCRFGG 20

RESULT 11

VFFDIT

phoratoxin - California mistletoe

C:Species: Phoradendron tomentosum (California mistletoe)

C>Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 15-Oct-1996

C:Accession: A01802

R:Melstrand, S.T.; Samuelsson, G.

Acta Pharm. Suec. 11, 347-360, 1974

A:Title: Phoratoxin, a toxic protein from the mistletoe Phoradendron tomentosusp. macro

A:Reference number: A90011; MUID:75013484; PMID:4415051

A:Note: P. t. subsp. macrophyllum

A:Accession: A01802

A:Molecule type: protein

A:Residues: 1-46 <MEL>

A:Note: the carboxyl end is blocked but is not amidated

R:Melstrand, S.T.; Samuelsson, G.

Acta Pharm. Suec. 11, 367-374, 1974

A:Title: Phoratoxin, a toxic protein from the mistletoe Phoradendron tomentosum subsp. m

A:Reference number: A90012; MUID:75013480; PMID:4606908

A:Contents: annotation; disulfide bonds

C:Comment: Phoratoxin, the single toxin of an American species of mistletoe, is related

C:Superfamily: viscotoxin

C:Keywords: blocked carboxyl end; toxin

F:3-40 4-32/Disulfide bonds: #status predicted

F:16-26/Disulfide bonds: #status experimental

F:46/Modified site: blocked carboxyl end (His) #status experimental

Query Match 0.9%; Score 7; DB 1; Length 46;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 NTCRFGG 18

|||||

Db 14 NTCRFGG 20

RESULT 12

S76642

hypothetical protein - Synecocystis sp. (strain PCC 6803)

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S76642

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shingo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76642

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A;Residues: 1-52 <KAN>
 A;Cross-references: EMBL:D64004; GB:AB001339; NID:g1001701; PIDN:BAA10586.1; PID:d101123
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C;Genetics:
 A;Start codon: GTG

Query Match 0.9%; Score 7; DB 2; Length 52;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 NIKRKTFF 169
 |||||
 DB 17 NIKRKTFF 23

RESULT 13
 AE1046
 SUGB protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain CT18)
 C;Species: Salmonella enterica subsp. enterica serovar typhi
 A;Note: this species has also been called Salmonella typhi
 C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C;Accession: AE1046
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
 A;Reference number: AB0502; MUID:21534947; PMID:11677608
 A;Accession: AE1046
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-105 <PAR>
 A;Cross-references: GB:AL513382; PIDN:CAD06818.1; PID:g16505468; GSPDB:GN00176
 C;Genetics:
 A;Gene: sugE
 C;Superfamily: sugE protein

Query Match 0.9%; Score 7; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 728 GESASPA 734
 |||||
 DB 79 GESASPA 85

RESULT 14
 F97132
 hypothetical protein CAC1885 [imported] - Clostridium acetobutylicum
 C;Species: Clostridium acetobutylicum
 C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C;Accession: F97132
 R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
 A;Reference number: A96900; MUID:21359325; PMID:21359325
 A;Accession: F97132
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-109 <KUR>
 A;Cross-references: GB:AB001437; PIDN:AAK79849.1; PID:g15024865; GSPDB:GN00168
 A;Experimental source: Clostridium acetobutylicum ATCC824
 C;Genetics:
 A;Gene: CAC1885

Query Match 0.9%; Score 7; DB 2; Length 109;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 734 ANEEITE 740
 |||||

Db 64 ANEEITE 70

RESULT 15

H90626
 NADH dehydrogenase chain 3 ND3 [imported] - Eudromia elegans mitochondrion
 C;Species: Mitochondrion Eudromia elegans
 C;Date: 15-Jun-2001 #sequence_revision 15-Jun-2001 #text_change 03-Aug-2001
 C;Accession: H90626
 R;Haddrath, O.; Baker, A.J.
 Proc. R. Soc. Lond. B Biol. Sci. 268, 939-945, 2001
 A;Title: Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetic
 A;Reference number: A99613; MUID:21263106; PMID:11370967
 A;Accession: H90626
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-116 <KUR>
 A;Cross-references: GB:NC_002772; NID:g14141811; PIDN:NP_115280.1; GSPDB:GN00163
 C;Genetics:
 A;Gene: ND3
 A;Genome: mitochondrion
 A;Genetic code: SGC1
 C;Superfamily: NADH dehydrogenase (ubiquinone) chain 3
 C;Keywords: mitochondrion

Query Match 0.9%; Score 7; DB 2; Length 116;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 LSLTLSS 90
 |||||
 DB 10 LSLTLSS 16

Search completed: August 25, 2004, 18:15:33
 Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 18:14:54 ; Search time 137 Seconds

(without alignments)
1782.041 Million cell updates/sec

Title: US-09-874-162A-8

Perfect score: 776

Sequence: 1 MTGIAAASFNTCRFGGCG.....KALETDSVSGVSKQKQKL 776

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1297172 seqs, 314612898 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1297172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:**

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----------------------|--------------------|
| 1 | 776 | 100.0 | 776 | US-09-874-162A-8 | Sequence 8, Appli |
| 2 | 747 | 83.4 | 739 | US-09-874-162A-5 | Sequence 5, Appli |
| 3 | 283 | 36.5 | 289 | US-09-764-864-911 | Sequence 911, App |
| 4 | 186 | 24.0 | 388 | US-09-764-864-931 | Sequence 931, App |
| 5 | 161 | 20.7 | 292 | US-09-764-864-1369 | Sequence 1369, Ap |
| 6 | 131 | 16.9 | 175 | US-09-764-864-1355 | Sequence 1355, Ap |
| 7 | 128 | 16.5 | 243 | US-09-874-162A-2 | Sequence 2, Appli |
| 8 | 128 | 16.5 | 243 | US-10-144-198-16 | Sequence 16, Appli |
| 9 | 128 | 16.5 | 243 | US-10-104-047-2404 | Sequence 2404, Ap |
| 10 | 109 | 14.0 | 278 | US-09-764-864-1334 | Sequence 1334, Ap |
| 11 | 93 | 12.0 | 143 | US-09-764-864-884 | Sequence 884, App |
| 12 | 67 | 8.6 | 69 | US-10-029-386-28495 | Sequence 28495, A |
| 13 | 56 | 7.2 | 171 | US-10-276-774-2060 | Sequence 2060, Ap |
| 14 | 8 | 1.0 | 25 | US-09-864-761-42997 | Sequence 42997, A |
| 15 | 8 | 1.0 | 150 | US-10-437-963-122951 | Sequence 122951, |

| | | | | | | |
|----|---|-----|------|----|----------------------|--------------------|
| 16 | 8 | 1.0 | 150 | 16 | US-10-437-963-147736 | Sequence 147736, |
| 17 | 8 | 1.0 | 151 | 12 | US-10-424-599-170262 | Sequence 170262, |
| 18 | 8 | 1.0 | 158 | 12 | US-10-424-599-170263 | Sequence 170263, |
| 19 | 8 | 1.0 | 283 | 12 | US-10-425-114-71700 | Sequence 71700, A |
| 20 | 8 | 1.0 | 285 | 12 | US-10-425-114-51976 | Sequence 51976, A |
| 21 | 8 | 1.0 | 297 | 9 | US-09-738-626-5528 | Sequence 5528, Ap |
| 22 | 8 | 1.0 | 324 | 12 | US-10-425-114-37645 | Sequence 37645, A |
| 23 | 8 | 1.0 | 372 | 12 | US-10-424-599-260059 | Sequence 260059, |
| 24 | 8 | 1.0 | 573 | 16 | US-10-287-092-34 | Sequence 34, Appli |
| 25 | 8 | 1.0 | 662 | 9 | US-09-924-358-11 | Sequence 11, Appli |
| 26 | 8 | 1.0 | 662 | 15 | US-10-410-764-11 | Sequence 7, Appli |
| 27 | 8 | 1.0 | 669 | 10 | US-09-790-853-7 | Sequence 123277, |
| 28 | 8 | 1.0 | 673 | 16 | US-10-437-963-123277 | Sequence 123277, |
| 29 | 8 | 1.0 | 1022 | 14 | US-10-301-822-163 | Sequence 163, App |
| 30 | 8 | 1.0 | 1022 | 16 | US-10-648-593-244 | Sequence 244, App |
| 31 | 8 | 1.0 | 1093 | 16 | US-10-275-998-1 | Sequence 1, Appli |
| 32 | 7 | 0.9 | 10 | 10 | US-09-572-404B-3143 | Sequence 3143, Ap |
| 33 | 7 | 0.9 | 10 | 10 | US-09-572-404B-3145 | Sequence 3145, Ap |
| 34 | 7 | 0.9 | 51 | 16 | US-10-437-963-162817 | Sequence 162817, |
| 35 | 7 | 0.9 | 54 | 16 | US-10-437-963-161985 | Sequence 161985, |
| 36 | 7 | 0.9 | 59 | 12 | US-10-424-599-163032 | Sequence 163032, |
| 37 | 7 | 0.9 | 60 | 10 | US-09-764-891-3469 | Sequence 3469, Ap |
| 38 | 7 | 0.9 | 63 | 12 | US-10-424-599-155981 | Sequence 155981, |
| 39 | 7 | 0.9 | 66 | 12 | US-10-424-599-259111 | Sequence 259111, |
| 40 | 7 | 0.9 | 69 | 16 | US-10-437-963-135864 | Sequence 135864, |
| 41 | 7 | 0.9 | 70 | 16 | US-10-437-963-147556 | Sequence 147556, |
| 42 | 7 | 0.9 | 74 | 12 | US-10-450-875-6 | Sequence 6, Appli |
| 43 | 7 | 0.9 | 89 | 12 | US-10-424-599-180215 | Sequence 180215, |
| 44 | 7 | 0.9 | 90 | 12 | US-10-424-599-255393 | Sequence 255393, |
| 45 | 7 | 0.9 | 94 | 10 | US-09-764-872-312 | Sequence 312, App |

ALIGNMENTS

RESULT 1
US-09-874-162A-8
; Sequence 8, Application US/09874162A
; Patent No. US20020155452A1
; GENERAL INFORMATION:
; APPLICANT: Koontz, Jason
; APPLICANT: Sklar, Jeffrey
; TITLE OF INVENTION: FUSION OF JAZF1 AND JJA21 GENES IN
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS
; FILE REFERENCE: 05311-024001
; CURRENT APPLICATION NUMBER: US/09/874,162A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,093
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-162A-8

| | | | | |
|-----------------------|---------|--|-------|-------------|
| Query Match | 100.0%; | Score 776; | DB 9; | Length 776; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches | 776; | Conservative | 0; | Mismatches |
| | | | 0; | Indels |
| | | | | Gaps |
| | | | | 0; |
| Qy | 1 | MTGIAAASFNTCRFGGCGLHPTTADLIEHEDNHDTPRVLEKQELQQPTTVALSY | 60 | |
| Db | 1 | MTGIAAASFNTCRFGGCGLHPTTADLIEHEDNHDTPRVLEKQELQQPTTVALSY | 60 | |
| Qy | 61 | INFMFMDAAREOESLKKIQPKLSLTLSSVSRGVNSTPPRHSSGSLTPPVPTPTPSS | 120 | |
| Db | 61 | INFMFMDAAREOESLKKIQPKLSLTLSSVSRGVNSTPPRHSSGSLTPPVPTPTPSS | 120 | |
| Qy | 121 | SFSSSTPTPTQYRFLRTRNLNFIPLHRTLTYMSHRNSTNKKRTFKVDMLSKVEK | 180 | |
| Db | 121 | SFSSSTPTPTQYRFLRTRNLNFIPLHRTLTYMSHRNSTNKKRTFKVDMLSKVEK | 180 | |

181 MKGEQESHLSAHLQLTFTGFFHKNDKPSNSENEQNSVTLVLLVKVCHKKRDXVSCPI 240
Db
181 MKGEQESHLSAHLQLTFTGFFHKNDKPSNSENEQNSVTLVLLVKVCHKKRDXVSCPI 240
Qy
241 ROVPTGKKQVPLIPDLNOKGPNFSLAVSSNEPEPNSHVMKYSYLLFRVTRPGRREFN 300
Db
241 ROVPTGKKQVPLIPDLNOKGPNFSLAVSSNEPEPNSHVMKYSYLLFRVTRPGRREFN 300
Qy
301 GMINGETNENIDVNEELPARKRNRDEGKTFVAQMTVFDKNRLLQLLDGEYEVAMQEME 360
Db
301 GMINGETNENIDVNEELPARKRNRDEGKTFVAQMTVFDKNRLLQLLDGEYEVAMQEME 360
Qy
361 ECPISKGAATWETILDGKRLPPFTFSOGPTLQFTLRWTGTNDKSTAPIAKPLATRNSE 420
Db
361 ECPISKGAATWETILDGKRLPPFTFSOGPTLQFTLRWTGTNDKSTAPIAKPLATRNSE 420
Qy
421 SLHQENKPGSVKPTOTIAVKSLETTDLQTRKEKDTPENRQKLRIFYQFLYNNNTROOTE 480
Db
421 SLHQENKPGSVKPTOTIAVKSLETTDLQTRKEKDTPENRQKLRIFYQFLYNNNTROOTE 480
Qy
481 ARDDLHCPCWTLNCRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKCYDGSYAGNP 540
Db
481 ARDDLHCPCWTLNCRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKCYDGSYAGNP 540
Qy
541 QDIHQPGFAPSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVEQOITYSSGHN 600
Db
541 QDIHQPGFAPSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVEQOITYSSGHN 600
Qy
601 LYFHSDDTCLPLRPOMEVDSDEKDPWLREKTTIQIEEFSDVNEGEKEVWKLWNLHVMK 660
Db
601 LYFHSDDTCLPLRPOMEVDSDEKDPWLREKTTIQIEEFSDVNEGEKEVWKLWNLHVMK 660
Qy
661 HGFIADNOMHACMLFVNYGQKIKNLRCNFMHLVSMHDFNLISIMSDKAVTKLRE 720
Db
661 HGFIADNOMHACMLFVNYGQKIKNLRCNFMHLVSMHDFNLISIMSDKAVTKLRE 720
Qy
721 MOOKLEKESASPAEETESQNGTANGFSEINSKEKALETDSVSGVSKSKQKL 776
Db
721 MOOKLEKESASPAEETESQNGTANGFSEINSKEKALETDSVSGVSKSKQKL 776

RESULT 2

US-09-874-162A-5

; Sequence 5, Application US/09874162A

; Patent No. US2002015542A1

; GENERAL INFORMATION:

; APPLICANT: Koontz, Jason

; APPLICANT: Sklar, Jeffrey

; TITLE OF INVENTION: FUSION OF JAZF1 AND JAZA1 GENES IN

; FILE REFERENCE: ENDOMETRIAL STROMAL TUMORS

; CURRENT APPLICATION NUMBER: US/09/874,162A

; CURRENT FILING DATE: 2001-06-04

; PRIOR APPLICATION NUMBER: US 60/209,093

; PRIOR FILING DATE: 2000-06-02

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 739

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-874-162A-5

Query Match

Best Local Similarity 83.4%; Score 647; DB 9; Length 739;

Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 PTQYRFLRTRNLIAPIFLHRTLYMHSRNRRTIKKRTFKVDDMLSKVEMKGEQESH 189

Db 93 PTQYRFLRTRNLIAPIFLHRTLYMHSRNRRTIKKRTFKVDDMLSKVEMKGEQESH 152

Qy 190 LSAHLQLTFTGFFHKNDKPSNSENEQNSVTLVLLVKVCHKKRDXVSCPIROVPTGKKQ 249

153 LSAHLQLTFTGFFHKNDKPSNSENEQNSVTLVLLVKVCHKKRDXVSCPIROVPTGKKQ 212
Qy
250 VPLIPDLNOKGPNFSLAVSSNEPEPNSHVMKYSYLLFRVTRPGRREFNMGINGETNE 309
Db
213 VPLIPDLNOKGPNFSLAVSSNEPEPNSHVMKYSYLLFRVTRPGRREFNMGINGETNE 272
Qy
310 NIDVNEELPARKRNRDEGKTFVAQMTVFDKNRLLQLLDGEYEVAMQEMECPI SKKRA 369
Db
273 NIDVNEELPARKRNRDEGKTFVAQMTVFDKNRLLQLLDGEYEVAMQEMECPI SKKRA 332
Qy
370 TWETILDGKRLPPFTFSOGPTLQFTLRWTGTNDKSTAPIAKPLATRNSSLHOENKPG 429
Db
333 TWETILDGKRLPPFTFSOGPTLQFTLRWTGTNDKSTAPIAKPLATRNSSLHOENKPG 392
Qy
430 SVKPTQTTIAVKSLETTDLQTRKEKDTPENRQKLRIFYQFLYNNNTROOTEARDDLHCPCW 489
Db
393 SVKPTQTTIAVKSLETTDLQTRKEKDTPENRQKLRIFYQFLYNNNTROOTEARDDLHCPCW 452
Qy
490 CTLNCRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKCYDGSYAGNPDIHQPGF 549
Db
453 CTLNCRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKCYDGSYAGNPDIHQPGF 512
Qy
550 AFSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVEQOITYSSGHNRLYFHSDDTCL 609
Db
513 AFSRNGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVEQOITYSSGHNRLYFHSDDTCL 572
Qy
610 PLRPOMEVDSDEKDPWLREKTTIQIEEFSDVNEGEKEVWKLWNLHVMKHGFIADNOM 669
Db
573 PLRPOMEVDSDEKDPWLREKTTIQIEEFSDVNEGEKEVWKLWNLHVMKHGFIADNOM 632
Qy
670 NEACMLFVNYGQKIKNLRCNFMHLVSMHDFNLISIMSDKAVTKLREMOOKLEKGE 729
Db
633 NEACMLFVNYGQKIKNLRCNFMHLVSMHDFNLISIMSDKAVTKLREMOOKLEKGE 692
Qy
730 SASPANEETESQNGTANGFSEINSKEKALETDSVSGVSKSKQKL 776
Db
693 SASPANEETESQNGTANGFSEINSKEKALETDSVSGVSKSKQKL 739

RESULT 3

US-09-764-864-911

; Sequence 911, Application US/09764864

; Patent No. US20020132753A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT223

; CURRENT APPLICATION NUMBER: US/09/764,864

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1792

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 911

; LENGTH: 289

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-864-911

Query Match

Best Local Similarity 36.5%; Score 283; DB 9; Length 289;

Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 494 CRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKCYDGSYAGNPDIHQPGFAFSR 553

Db 7 CRKLYSLKHLKCHSRFIYNYVHPKGARIDVSIKCYDGSYAGNPDIHQPGFAFSR 66

Qy 554 NGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVEQOITYSSGHNRLYFHSDDTCLPLRP 613

Db 67 NGPVKRTPTIHLVCRPKRTKASMEFLESEDEGEVEQOITYSSGHNRLYFHSDDTCLPLRP 126

Qy 614 QMEVDSDEKDPWLREKTTIQIEEFSDVNEGEKEVWKLWNLHVMKHGFIADNOMHAC 673

Db 127 QMEVDSDEKDPWLREKTTIQIEEFSDVNEGEKEVWKLWNLHVMKHGFIADNOMHAC 186

```
QY 674 MLFVNYGQKIIKKNLCRNFMHLVSMHDFNLSIMSDKAVTKLRMOQKLEKESASP 733
Db 187 MLFVNYGQKIIKKNLCRNFMHLVSMHDFNLSIMSDKAVTKLRMOQKLEKESASP 246
QY 734 ANEEITEONGTANGFSEINSKEKALETDSVSGSKQKOKL 776
Db 247 ANEEITEONGTANGFSEINSKEKALETDSVSGSKQKOKL 289

RESULT 4
US-09-764-864-931
; Sequence 931, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 931
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-931

Query Match 24.0%; Score 186; DB 9; Length 388;
Best Local Similarity 99.5%; Pred. No. 1.6e-174;
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 389 GPTLOFTLWGTGETNDKSTAPIAKPLATRNSSLHQENKPGSVKPTQIAVKESLTTDLQ 448
Db 1 GPTLOFTLWGTGETNDKSTAPIAKPLATRNSSLHQENKPGSVKPTQIAVKESLTTDLQ 60
QY 449 TRKEKDTPNENRQKRIFYQFLYNNNTROOTEARDLHCPWCTLNCRLKLYSLHLKLCH 508
Db 61 TRKEKDTPNENRQKRIFYQFLYNNNTROOTEARDLHCPWCTLNCRLKLYSLHLKLCH 120
QY 509 SRPIFYVYHPKGARIDVINECYGSGVAGNPQDIHQPGFAPSRNGPVKPTPIHILVC 568
Db 121 SRPIFYVYHPKGARIDVINECYGSGVAGNPQDIHQPGFAPSRNGPVKPTPIHILVC 180
QY 569 RPKRTKASSEFLESEDEGEVEQRTYSSGHNLRYFHSHTCLPLRQEMEVDSDEKDPW 628
Db 181 RPKRTKASSEFLESEDEGEVEQRTYSSGHNLRYFHSHTCLPLRQEMEVDSDEKDPW 240
QY 629 LREKTIQTIEEFDVNEGEKEYMKLWNLHVMKHGFIADNQMNHACMLFVNYGQKIIKKN 688
Db 241 LREKTIQTIEEFDVNEGEKEYMKLWNLHVMKHGFIADNQMNHACMLFVNYGQKIIKKN 300
QY 689 LCRNFMHLVSMHDFNLSIMSDKAVTKLRMOQKLEKESASPANEEITEONGTANG 748
Db 301 LCRNFMHLVSMHDFNLSIMSDKAVTKLRMOQKLEKESASPANEEITEONGTANG 360
QY 749 FSEINSKEKALETDSVSGSKQKOKL 776
Db 361 FSEINSKEKALETDSVSGSKQKOKL 388

RESULT 5
US-09-764-864-1369
; Sequence 1369, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
```

```
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1369
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (9)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (31)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (46)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (121)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (283)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1369

Query Match 20.7%; Score 161; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 5.9e-150;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 VAMQEMECPISKGRATWETILDGKRLPPETFSQGSTLTQTLRWGTETNDKSTAPIAKP 413
Db 122 VAMQEMECPISKGRATWETILDGKRLPPETFSQGSTLTQTLRWGTETNDKSTAPIAKP 181
QY 414 LATRNSSLHQENKPGSVKPTQIAVKESLTTDLQTRKEKDTPNENRQKRIFYQFLYNN 473
Db 182 LATRNSSLHQENKPGSVKPTQIAVKESLTTDLQTRKEKDTPNENRQKRIFYQFLYNN 241
QY 474 NTRQOTEARDLHCPWCTLNCRLKLYSLHLKLCHSRFIEN 514
Db 242 NTRQOTEARDLHCPWCTLNCRLKLYSLHLKLCHSRFIEN 282

RESULT 6
US-09-764-864-1355
; Sequence 1355, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1355
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (166)
```

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1355

Query Match 16.5%; Score 131; DB 9; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.9e-120; Indels 0; Gaps 0;
Matches 131; Conservative 0; Mismatches 0;
QY 470 LYNNTRQQTAEADLDHCPWCTLNCRKLYSLKHLKCHSRFIFNYVHPKGARIDVSN 529
DB 5 LYNNTRQQTAEADLDHCPWCTLNCRKLYSLKHLKCHSRFIFNYVHPKGARIDVSN 64
QY 530 EYDGSYAGNPODIHQGFASFSGNPGVKTPTITHILVCRPKTKASMEFESEDEGEVE 589
DB 65 EYDGSYAGNPODIHQGFASFSGNPGVKTPTITHILVCRPKTKASMEFESEDEGEVE 124
QY 590 QORTYSSGHNH 600
DB 125 QORTYSSGHNH 135

RESULT 7
US-09-874-162A-2
; Sequence 2, Application US/09874162A
; Patent No. US20020155452A1
; GENERAL INFORMATION:
; APPLICANT: Koonetz, Jason
; APPLICANT: Sklar, Jeffrey
; TITLE OF INVENTION: FUSION OF JAZF1 AND JUJAZ1 GENES IN
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS
; FILE REFERENCE: 05311-024001
; CURRENT APPLICATION NUMBER: US/09/874,162A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,093
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-162A-2

Query Match 16.5%; Score 128; DB 9; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.9e-117; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 0;
QY 1 MTGIAAASFNTCRFGGGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
DB 1 MTGIAAASFNTCRFGGGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
QY 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
DB 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
QY 121 SFRSSTPT 128
DB 121 SFRSSTPT 128

RESULT 8
US-10-144-198-16
; Sequence 16, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 243

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-16
Query Match 16.5%; Score 128; DB 15; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.9e-117; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 0;
QY 1 MTGIAAASFNTCRFGGGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
DB 1 MTGIAAASFNTCRFGGGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
QY 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
DB 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
QY 121 SFRSSTPT 128
DB 121 SFRSSTPT 128

RESULT 9
US-10-104-047-2404
; Sequence 2404, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2404
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2404

Query Match 16.5%; Score 128; DB 15; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.9e-117; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 0;
QY 1 MTGIAAASFNTCRFGGGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
DB 1 MTGIAAASFNTCRFGGGLHFPPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
QY 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
DB 61 INRFMTDAARREQESLKKIKPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
QY 121 SFRSSTPT 128
DB 121 SFRSSTPT 128

RESULT 10
US-09-764-864-1334
; Sequence 1334, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1334
; LENGTH: 278


```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1334

Query Match
Best Local Similarity 14.0%; Score 109; DB 9; Length 278;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHIEDNHIDTDPVLEKQLQOPTVALSY 60
Db 36 MTGIAAASFNTCRFGGGLHPTLADLIEHIEDNHIDTDPVLEKQLQOPTVALSY 95
QY 61 INRFMTDAARRQESLKKKIQKLSLTSSSVSRGNVSTPPRHSSGSLT 109
Db 96 INRFMTDAARRQESLKKKIQKLSLTSSSVSRGNVSTPPRHSSGSLT 144

RESULT 11
US-09-764-864-884
; Sequence 884, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 884
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-884

Query Match
Best Local Similarity 12.0%; Score 93; DB 9; Length 143;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHIEDNHIDTDPVLEKQLQOPTVALSY 60
Db 51 MTGIAAASFNTCRFGGGLHPTLADLIEHIEDNHIDTDPVLEKQLQOPTVALSY 110
QY 61 INRFMTDAARRQESLKKKIQKLSLTSSSVS 93
Db 111 INRFMTDAARRQESLKKKIQKLSLTSSSVS 143

RESULT 12
US-10-029-386-28495
; Sequence 28495, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28495
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; OTHER INFORMATION: MAP TO CHRI7.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P23477, EVALUATE 2.70e+00
US-10-029-386-28495

Query Match
Best Local Similarity 8.6%; Score 67; DB 14; Length 69;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 RPKETKASMEFFLESEDEGEVQORTYSSGHNRLYFHSDTCLPLRQSEYDSEDEKDPW 628
Db 1 RPKETKASMEFFLESEDEGEVQORTYSSGHNRLYFHSDTCLPLRQSEYDSEDEKDPW 60
QY 629 LREXITIT 635
Db 61 LREXITIT 67

RESULT 13
US-10-276-774-2060
; Sequence 2060, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2060
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-2060

Query Match
Best Local Similarity 7.2%; Score 56; DB 12; Length 171;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 QESLKKKIQPKLSLTSSSVSRGNVSTPPRHSSGSLTPPVTPITSSSPSSSTPT 128
Db 1 QESLKKKIQPKLSLTSSSVSRGNVSTPPRHSSGSLTPPVTPITSSSPSSSTPT 56

RESULT 14
US-09-864-761-42997
; Sequence 42997, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366

```

```
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42997
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012111.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
US-09-864-761-42997
```

```
Query Match 1.0%; Score 8; DB 9; Length 25;
Best Local Similarity 100.0%; Pred.No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 110 PPVTPPT 117
Db 13 PPVTPPT 20
```

```
RESULT 15
US-10-437-963-122951
; Sequence 122951, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
```

```
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122951
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MPT4530_25832C.1.pep
US-10-437-963-122951

Query Match 1.0%; Score 8; DB 16; Length 150;
Best Local Similarity 100.0%; Pred.No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 408 APIAKPLA 415
Db 17 APIAKPLA 24

Search completed: August 25, 2004, 18:26:48
Job time : 140 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 18:09:41 ; Search time 33 Seconds
(without alignments)
1213.993 Million cell updates/sec

Title: US-09-874-162A-8
Perfect score: 776
Sequence: 1 MTGIAAGFFNTCRFGCG.....KALETDSVSGVSKOSKQKL 776

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1 | 10 | 1.3 | 64 | US-09-543-681A-6424 | Sequence 6424, Ap |
| 2 | 9 | 1.2 | 96 | US-09-328-352-4313 | Sequence 4313, Ap |
| 3 | 7 | 0.9 | 139 | US-09-770-834-13 | Sequence 13, Appl |
| 4 | 7 | 0.9 | 142 | US-09-328-352-4320 | Sequence 4320, Ap |
| 5 | 7 | 0.9 | 151 | US-09-621-976-4980 | Sequence 4980, Ap |
| 6 | 7 | 0.9 | 155 | US-09-489-039A-14046 | Sequence 14046, A |
| 7 | 7 | 0.9 | 174 | US-09-328-352-7878 | Sequence 7878, Ap |
| 8 | 7 | 0.9 | 178 | US-09-328-352-4296 | Sequence 4296, Ap |
| 9 | 7 | 0.9 | 196 | US-09-842-164A-2 | Sequence 2, Appli |
| 10 | 7 | 0.9 | 206 | US-09-252-991A-31591 | Sequence 31591, A |
| 11 | 7 | 0.9 | 247 | US-09-134-001C-5610 | Sequence 5610, Ap |
| 12 | 7 | 0.9 | 279 | US-09-107-532A-6807 | Sequence 6807, Ap |
| 13 | 7 | 0.9 | 294 | US-08-973-334-2 | Sequence 2, Appli |
| 14 | 7 | 0.9 | 294 | US-09-563-869A-2 | Sequence 2, Appli |
| 15 | 7 | 0.9 | 294 | US-08-549-489-2 | Sequence 2, Appli |
| 16 | 7 | 0.9 | 300 | US-09-252-991A-17565 | Sequence 17565, A |
| 17 | 7 | 0.9 | 320 | US-08-579-940-8 | Sequence 8, Appli |
| 18 | 7 | 0.9 | 324 | US-09-010-147B-16 | Sequence 16, Appl |
| 19 | 7 | 0.9 | 332 | US-09-498-520A-10 | Sequence 10, Appl |
| 20 | 7 | 0.9 | 332 | US-09-134-000C-4988 | Sequence 4988, Ap |
| 21 | 7 | 0.9 | 333 | US-09-540-236-2603 | Sequence 2603, Ap |
| 22 | 7 | 0.9 | 368 | US-08-630-915A-20 | Sequence 20, Appl |
| 23 | 7 | 0.9 | 370 | US-09-107-532A-5268 | Sequence 5268, Ap |
| 24 | 7 | 0.9 | 384 | US-09-198-452A-1111 | Sequence 1111, Ap |
| 25 | 7 | 0.9 | 402 | US-09-721-870-46 | Sequence 46, Appl |
| 26 | 7 | 0.9 | 493 | US-09-252-991A-27576 | Sequence 27576, A |
| 27 | 7 | 0.9 | 519 | US-08-996-338-21 | Sequence 21, Appl |

| | | | | | | |
|----|---|-----|-----|---|----------------------|-------------------|
| 28 | 7 | 0.9 | 519 | 4 | US-09-556-972-21 | Sequence 21, Appl |
| 29 | 7 | 0.9 | 526 | 4 | US-09-198-452A-347 | Sequence 347, App |
| 30 | 7 | 0.9 | 534 | 3 | US-08-875-223-8 | Sequence 8, Appli |
| 31 | 7 | 0.9 | 537 | 1 | US-08-604-333-4 | Sequence 4, Appli |
| 32 | 7 | 0.9 | 537 | 3 | US-09-110-618-4 | Sequence 4, Appli |
| 33 | 7 | 0.9 | 537 | 4 | US-09-173-151A-29 | Sequence 29, Appl |
| 34 | 7 | 0.9 | 537 | 4 | US-09-578-178-4 | Sequence 4, Appli |
| 35 | 7 | 0.9 | 537 | 4 | US-09-577-806-4 | Sequence 4, Appli |
| 36 | 7 | 0.9 | 537 | 4 | US-09-621-502-8 | Sequence 8, Appli |
| 37 | 7 | 0.9 | 566 | 3 | US-09-073-297-5 | Sequence 5, Appli |
| 38 | 7 | 0.9 | 570 | 3 | US-08-991-944-2 | Sequence 2, Appli |
| 39 | 7 | 0.9 | 570 | 3 | US-08-991-944-4 | Sequence 4, Appli |
| 40 | 7 | 0.9 | 570 | 4 | US-09-173-151A-21 | Sequence 21, Appl |
| 41 | 7 | 0.9 | 570 | 4 | US-09-173-151A-36 | Sequence 36, Appl |
| 42 | 7 | 0.9 | 602 | 4 | US-09-489-039A-7207 | Sequence 7207, Ap |
| 43 | 7 | 0.9 | 690 | 4 | US-09-252-991A-32350 | Sequence 32350, A |
| 44 | 7 | 0.9 | 887 | 4 | US-09-077-940A-2 | Sequence 2, Appli |
| 45 | 7 | 0.9 | 888 | 4 | US-09-077-940A-4 | Sequence 4, Appli |

ALIGNMENTS

RESULT 1
US-09-543-681A-6424
; Sequence 6424, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6424
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6424

Query Match 1.3%; Score 10; DB 4; Length 64;
Best Local Similarity 100.0%; Pred. No. 0.029; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

| | | | |
|----|----|------------|----|
| Qy | 85 | SLTLSSSVSR | 94 |
| | | | |
| Db | 48 | SLTLSSSVSR | 57 |

RESULT 2
US-09-328-352-4313
; Sequence 4313, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4313
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4313

Query Match 1.2%; Score 9; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 0.44; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 109 TPPVTPPIT 117
 |||||
 Db 55 TPPVTPPIT 63

RESULT 3

US-09-770-834-13
 ; Sequence 13, Application US/09770834
 ; Patent No. 6684162

GENERAL INFORMATION:

; APPLICANT: Parris, Kevin
 ; APPLICANT: Somers, William
 ; APPLICANT: Tam, Amy
 ; APPLICANT: Lin, Laura
 ; APPLICANT: Stahl, Mark
 ; APPLICANT: Powers, Robert
 ; APPLICANT: Xu, Guan-Yi
 ; TITLE OF INVENTION: CRYSTAL STRUCTURE OF ACPS/ACP COMPLEX, SOLUTION STRUCTURE
 ; FILE OF INVENTION: OF B. SUBTILIS ACP, AND USES THEREOF
 ; FILE REFERENCE: 2368/14
 ; CURRENT APPLICATION NUMBER: US/09/770,834
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: US 60/202,466
 ; PRIOR FILING DATE: 2000-05-08
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: Patent version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 139
 ; TYPE: PRT
 ; ORGANISM: Bradyrhizobium sp.

US-09-770-834-13

Query Match 0.9%; Score 7; DB 4; Length 139;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 522 ARIDVSI 528
 |||||
 Db 112 ARIDVSI 118

RESULT 4

US-09-328-352-4320
 ; Sequence 4320, Application US/09328352
 ; Patent No. 6562958

GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 4320
 ; LENGTH: 142
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii

US-09-328-352-4320

Query Match 0.9%; Score 7; DB 4; Length 142;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 PPVTPPI 116
 |||||
 Db 37 PPVTPPI 43

RESULT 5

US-09-621-976-4980
 ; Sequence 4980, Application US/09621976
 ; Patent No. 6639063

GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Jobert, S.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 ; FILE REFERENCE: GENSET.054PR2
 ; CURRENT APPLICATION NUMBER: US/09/621,976
 ; CURRENT FILING DATE: 2000-07-21
 ; NUMBER OF SEQ ID NOS: 19335
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 4980
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-621-976-4980

Query Match 0.9%; Score 7; DB 4; Length 151;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 INECYDG 534
 |||||
 Db 55 INECYDG 61

RESULT 6

US-09-489-039A-14046
 ; Sequence 14046, Application US/09489039A
 ; Patent No. 6610836

GENERAL INFORMATION:

; APPLICANT: Gary Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 14046
 ; LENGTH: 155
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-14046

Query Match 0.9%; Score 7; DB 4; Length 155;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 521 GARIDVS 527
 |||||
 Db 57 GARIDVS 63

RESULT 7

US-09-328-352-7878
 ; Sequence 7878, Application US/09328352
 ; Patent No. 6562958

GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 7878
 ; LENGTH: 174
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii

US-09-328-352-7878

Query Match 0.9%; Score 7; DB 4; Length 174;

```
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 584 EDGEVEQ 590
Db 57 EDGEVEQ 63

RESULT 8
US-09-328-352-4296
; Sequence 4296, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4296
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4296

Query Match 0.9%; Score 7; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 PPTTPI 116
Db 138 PPTTPI 144

RESULT 9
US-09-842-164A-2
; Sequence 2, Application US/09842164A
; Patent No. 6544754
; GENERAL INFORMATION:
; APPLICANT: INOUE, SATOSHI
; TITLE OF INVENTION: LUCIFERASE AND PHOTOPROTEIN
; FILE REFERENCE: 206497US0
; CURRENT APPLICATION NUMBER: US/09/842,164A
; CURRENT FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: JP 2000-125053
; PRIOR FILING DATE: 2000-04-26
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Oplophorus graciliorostris
US-09-842-164A-2

Query Match 0.9%; Score 7; DB 4; Length 196;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 SLLFRVT 292
Db 175 SLLFRVT 181

RESULT 10
US-09-252-991A-31591
; Sequence 31591, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
```

```
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31591
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31591

Query Match 0.9%; Score 7; DB 4; Length 206;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 SLSAHLQ 195
Db 91 SLSAHLQ 97

RESULT 11
US-09-134-001C-5610
; Sequence 5610, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5610
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5610

Query Match 0.9%; Score 7; DB 4; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 500 LLXHLKL 506
Db 124 LLXHLKL 130

RESULT 12
US-09-107-532A-6807
; Sequence 6807, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
```



```

;
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-563-869A-2

Query Match          0.9%; Score 7; DB 3; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      379 RLPPPET 385
Db      29 RLPPPET 35

RESULT 15
US-08-549-489-2
; Sequence 2, Application US/08549489
; Patent No. 6281010
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Gao, Guang-Ping
; TITLE OF INVENTION: No. 6281010el Adenovirus Gene Therapy Vehicle
; TITLE OF INVENTION: and Cell Line
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Howson and Howson
; STREET: Box 457, 321 No. 6281010ristown Road
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,489
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/462,014
; FILING DATE: 08-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVFN013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 540-9206
; TELEFAX: (215) 540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-549-489-2

Query Match          0.9%; Score 7; DB 3; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      379 RLPPPET 385
Db      29 RLPPPET 35

Search completed: August 25, 2004, 18:16:11
Job time : 35 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:43:38 ; Search time 133 Seconds
(without alignments)
1840.918 Million cell updates/sec

Title: US-09-874-162A-8
Perfect score: 4078
Sequence: 1 MTGIAAASFNTCRFGCG.....KALETDSVSGVSKQKQL 776

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriaph.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 244.5 | 6.0 | 626 | 10 Q84UX5 | Q84UX5 zea mays (m |
| 2 | 224 | 5.5 | 628 | 5 Q9VZ26 | Q9VZ26 drosophila |
| 3 | 205.5 | 5.0 | 614 | 10 Q7XSM1 | Q7XSM1 oryza sativ |
| 4 | 177 | 4.3 | 1422 | 5 Q8IEL9 | Q8IEL9 plasmodium |
| 5 | 174 | 4.3 | 1693 | 5 Q86H41 | Q86H41 dictyosteli |
| 6 | 170.5 | 4.2 | 1396 | 11 Q810T3 | Q810T3 mus musculu |
| 7 | 166 | 4.1 | 1507 | 6 Q95151 | Q95151 canis fami |
| 8 | 163 | 4.0 | 1314 | 3 Q53904 | Q53904 kluyveromyc |
| 9 | 160.5 | 3.9 | 2847 | 5 Q9U4X0 | Q9U4X0 plasmodium |
| 10 | 159.5 | 3.9 | 1401 | 6 Q8HY4 | Q8HY4 bos taurus |
| 11 | 158 | 3.9 | 3471 | 6 Q8MK19 | Q8MK19 canis fami |
| 12 | 157.5 | 3.9 | 5767 | 5 Q81525 | Q81525 plasmodium |
| 13 | 157 | 3.8 | 1011 | 2 Q9AHL0 | Q9AHL0 borrelia bu |
| 14 | 157 | 3.8 | 1163 | 5 Q81KP5 | Q81KP5 plasmodium |
| 15 | 157 | 3.8 | 1893 | 5 Q9W4G6 | Q9W4G6 drosophila |
| 16 | 156.5 | 3.8 | 1065 | 2 Q9AHK7 | Q9AHK7 borrelia bu |

ALIGNMENTS

RESULT 1

Q84UX5 Q84UX5 PRELIMINARY; PRT; 626 AA.
ID AC Q84UX5
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE VEF family protein.
GN VEF101.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. B73;
RA Chandler V.L., Kaeppler S.M., Kaeppler H.F., Cone K.C.;
RT "Sequences from the Plant Chromatin Consortium (NSF Plant Genome
program Grant 9975930).";
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. B73;
RA Bergstrom D.E., Springer N.M., Schmitt L., Guthrie E., Sidorenko L.,
RA Selinger D., Kaeppler S.M., Cone K.C.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY232824; AAC04022.1; -
DR InterPro; IPR007087; Znf C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ SEQUENCE 626 AA; 71401 MW; 604C0D46B5A8718 CRC64;

Query Match 6.0%; Score 244.5; DB 10; Length 626;
Best Local Similarity 18.1%; Pred. No. 3.7e-09;
Matches 123; Conservative 110; Mismatches 251; Indels 197; Gaps 22;
Qy 129 EPTQIVREFLTRNLAPIFLHRTLTVMHSRNSRNTNKRKTFKVDMLSKVKMKGEQESH 188
Db 53 KPVELYNIORRAWKNPFIQRCLLYNIHARRKRIQ----- 89

[illegible]


```

Db      655 KCRLL-ILQNNIEMITDIIVFEYTESYRNTENTENEGNOCTDAGRNTCNSESOGSDSSKNDTV 713
QY      312 DVNEE---LPARRKENREDGKFTVAQMTVPDKRRRLQLLDGEYEV---AMQMEECPI 364
Db      714 YIHEENGSLPCIDQHNID-----LKLPQFMKEGNTQIQEGLSDLTCLVEMKAEETSHV 767
QY      365 SKKRATWETILDGKLPFPFEP---SQGPTLQFTLRWGTETNDKSTAPIAKPLATRN---- 418
Db      768 TWSNKKQQTANTGQNIKDFDYFYLSPFQASRKNIIRVSLSLNKARSLLNOKWTEELNPF 827
QY      419 SESLHQENKPG-SVKPTQTIIVAKESLTLTDLQTRKEKTPNENRQKLRIFQYFLYNNNTRQ 477
Db      828 SDLSNELLPGLDIKKT-----ISNHEVIENTERDKITKESDLGLGTENILLILOORPESK 884
QY      478 QTEARDDHLCPWCTLNCRLKLYSLKHLKCHSRP-----IFNVYHPKGRIDVSI 528
Db      885 IKKIKESAVLGFHTASGKKIETIKESLSDKVNKLFEKEQDNSEITNFESH--RGAKMSKDR 942
QY      529 NECYDG-SVAGNPDIHQPGFAPSRNGPVRKTPITH-ILVCRP-----KRT----- 573
Db      943 BECKDGRACGTETITTPYEETHSSLEKLVNSHIAALRPRLSLDNLKOTENLKI 1002
QY      574 -----KASMEFELESDGE-----VEQRTYSSGHNRLYFPHSDTCLPLRPQMEVDSED 622
Db      1003 SDHASQKVDVHENTEKETAKKPTMYTNQSTYSATENSPLTFTQDT-----EEKFSVSEA 1056
QY      623 E--KDPEWLRKTIQIIEPDS-----VNEGEKEVMKLM-----NLHVWKHGFADNQ 668
Db      1057 SLFEAKKWLRE-----GEWDQSERINAANKVCLKYPDYDPENPCGNSNSAITEND 1110
QY      669 MNHACMLFVENYGOKIIKKNLCRN-FMLHLVSMHD---FNLISIMS-----IDKAVT 716
Db      1111 KNH-----LSEKQSYLSNWSMSYSVHPGFCHSESVYNKSEVLSKIDNSGIEBPVIK 1166
QY      717 KLEMQ-----OKLEKGESASPAN----- 735
Db      1167 NIREKKNIGFSEIMSPGREADTPQSVNEDICVEKLTATNSCKKNKTAIKVAISDSNFFN 1226
QY      736 --EETEQNGTANGFSBINSKE-----KALE--TDSVSGVSKQKQK 775
Db      1227 TIQKLSNDSNNSVPAYSIVNSKRVFAHQTKEGFTDNCNCSWTKQNTKSK 1277

RESULT 8
ID O59904 PRELIMINARY; PRT; 1314 AA.
AC O59904;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Silent information regulator 4.
GN SIR4.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98198828; PubMed=9539421;
RA Astrom S.U., Rine J.;
RT "Theme and variation among silencing proteins in Saccharomyces
RL cerevisiae and Kluyveromyces lactis.";
RL Genetics 148:1021-1029 (1998).
DR EMBL; AF035007; AAC39438.1; -.
DR PIR; T09481; T09481.
SQ SEQUENCE 1314 AA; 146735 MW; 3C78CBB63319F8A6 CRC64;

Query Match 4.0%; Score 163; DB 3; Length 1314;
Best local similarity 19.9%; Pred. No. 0.008;
Matches 169; Conservative 140; Mismatches 338; Indels 204; Gaps 39;
QY 43 RVLEKQEL--QOPTYVALSYINRFTMDARREQBSLKK---KIQPKLSLTLSVSVS--RG 95

```

```

Db      181 RVVEDEVLRQOSSUSSGRHGHIESFQRMKPVSKSVKAPKNVAFAFETSUNDIDL 240
QY      96 NVSTPPRHSSGLTPPVTPPTPSSFRSSTPTEPTQIYRFLTRN-----LIAPFL 148
Db      241 NLLTRARASSGG-----TPTSGSAAPVAP-----KTCKNTDSQVSKIANBIQ 284
QY      149 HRTLYMHRNRTWIKKTFKVDMDLSKVKKMKGEQESHLSAHLQLTFTGFFHKNDKP 208
Db      285 NENKGRISKANS--PIQRTTTTTSNIPSEAKEL-----YNMLSSRLKRNVLFDKKGTTN 337
QY      209 SPNSNEQNSVTLEVLVKKVCHKKGVSCPIRQVPTKCKQVPLPDLNQTQKPGFPPLA 268
Db      338 SENNDKTGASVKVESHVVEILGSKEDHTHSFELPSPMQATIADLAKTITSLVKENSAQIE 397
QY      269 VSSNFEPSNSHMVKSYSLLFRVTPRRRPFNGMGTENENIDVNEEL-----PARR 321
Db      398 VDRNLIGPERS-----RKNKEFQDL---ETRTSSPAKESLASPSADTPERK 441
QY      322 KENREDGKTFVAQMTVPDKRRRLQLLDGEY-----VAMQMEECPI8----- 365
Db      442 VPQLDLTSPSPSTPGVTEKMDLKEVPSPSPSSAGAVQOSESQISHVITQAGADSG 501
QY      366 -----KKRATWETILDGKLPFPFETFSQGTPLQFTLRWGTETNDKSTAPIAKPLATR 417
Db      502 SHHGQIVESSELNIPTEPKDN-----QAASASGTFHHVDE--KOSTBETAKLLALS 553
QY      418 NESLHQENKPGS-----VKPTQTIIVAKESLTLTDLQTRKEKTPNENRQKLRIFQYFLN 472
Db      554 DTSS--ETSSGSSSGHYVGDQFLLNTWKARWGPVAKWAP---NKSFPDLMKLYLIA 608
QY      473 NNTRQOTEARDDLHCPWCTLNCRLKLYSL-LKHLKCHSRFIF-----NYVYHPKGA- 522
Db      609 EHT--PLAARSDVSCNDICNIRNYESVEVHLQKLPRLAYLELESPPSNFISKPLRAN 666
QY      523 -RID-VSINECDGSGYAGNPDIHQPGFAPSRNGPVK-----RTPIT 563
Db      667 NRVEKLRKDELNGS-----KQKRTGSPDSESGKKPLIERQIEDGNVSEKTPQEPPT 721
QY      564 HLIVCRPKRTKASMEFELESDGEVEQRTYSSGHNRLYFPHSDT---CLP----- 610
Db      722 SI-----PGETDVNEKVLPIED--IEMLPFTFISSNRLRLEPDANTLSAHLFSSVIGVKQ 775
QY      611 -----LRPQMEVDSE---DEKDEWLRKXTITQIEFSDVNEGEKVM 651
Db      776 PSLLTVNKPFPASFENGLOPDNAQASTEIETNEGNSKGGKDELHFDPERCEDVTQFRKEYL 835
QY      652 KLNHLVWKHGFADNQNHACMLFVENYGOKIIKKNLCRNFMHLVSMHDFNLISIMSI 711
Db      836 AL-RASISKDDTCSSSLN-----EDLNKTEVIQNLTSIIL--QSEWKIASLVGI 882
QY      712 DKAVTKLREMOKLEK--GESASPANEIITBEQN---GTANGFSRINSKEALETDSVS 765
Db      883 NH---QLREKLEBELEKIQAKLVQELDTVLLESKGSFSAGTAKA-AEVNKKDK--TESIN 935
QY      766 GYSKQSKQKQL 776
Db      936 --NEQAREKSL 944

RESULT 9
ID Q9U4X0 PRELIMINARY; PRT; 2647 AA.
AC Q9U4X0;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Putative erythrocyte binding protein EBL-1 (Fragment).
GN EBL-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]

```



```
QY 461 --OKLRIFQFLYNNTRQ---QTEARD-----DLHCPWCTLNCRLKLSL-LK 502
Db 1195 EVSKLQSIQ-----NTQALKKLETRVVDLSKYKATKSDELTQISDLN-EKANLNK 1238
QY 503 HLKCHSRFIRVYVHPKARIDVINECYGSGVAGNPQDIHQPGFAPSRNGVPVKTPPI 562
Db 1239 YEEVCEE-----VLHAK--KKELSAKDEKELLHFSIQEIKDQ-----QBRCDKSLTTI 1285
QY 563 THILVCRPKRKASMSFLESED-----GEVQQRYSNGHNLVYFHSDTCLPLRPOE 615
Db 1286 TEL-----QRRQESAKQ--TEAKONKITELNDVRLKQALNGLSOLTYGSGS---PSRQS 1338
QY 616 MEVDSSEKDPWMLREKTIITIEEFSDVNEGEKEVMKLNJHVM 659
Db 1339 QLIDS-----LQQQVRSLLQQQLADADRQHQEVIAIYRTHLL 1374

RESULT 11
Q8MKI9 PRELIMINARY; PRT; 3471 AA.
AC Q8MKI9
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Brca2 protein.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21570986; PubMed=11714026;
RA Ochiai K., Morimatsu M., Tomizawa N., Syuto B.;
RT "Cloning and Sequencing Full Length of Canine Brca2 and Rad51 cDNA.";
RL J. Vet. Med. Sci. 63:1103-1108 (2001).
DR EMBL; AB043895; BAB91245.2;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR00194; ATPase_a/bcentre.
DR Pfam; PF00634; BRCA2; 7.
DR PROSITE; PS00152; ATPASE ALPHA BETA; 1.
DR PROSITE; PS00138; BRCA2_REPEAT; 7.
SQ SEQUENCE 3471 AA; 389085 MW; 4AB77908A8EF6561 CRC64;

Query Match 3.9%; Score 158; DB 6; Length 3471;
Best Local Similarity 18.2%; Pred. No. 0.069;
Matches 177; Conservative 146; Mismatches 364; Indels 284; Gaps 44;

QY 23 PPTLADLIEHDNHDITDPRVLEKQELQOQTYVA--LSYINRFMTDAARQESLKKKIQ 81
Db 1023 FKTAGNKEIKUSNNIRKSKMLFKDIEHYPTNLACLEIVNTSLESQKPKSKS--HALD 1080
QY 82 PKLSLTLSSSVRGNVSTPPRHSSGSLPPVTPPTTSSPSRSTPEPTQI----- 133
Db 1081 PQSINIISGFVQ--NSTVSDSESGHTAPPTLSLKQPDFSNRNLTPSQKABITELSTILE 1138
QY 134 -----YRLRTNLIAPIFLHRTLTYMHSRNT--NIRKTFKVDK-----LSKV 178
Db 1139 ESGSQFETQPRK---PSHIIQRNPFEMPENQLNLNLSKWKXDDDLHLTTNAPSISQV 1195
QY 179 ERMK-----GEQESHLSLA-----HLQLTFTGFFH-KNDKPSNSEE----- 213
Db 1196 DSKKEGIGKQKQKACLSRLSCNRSASGYSTDKNEVEFRGFYSARGTKLVNVSALQKA 1255
QY 214 -----NEQNSV-----TLEVLVVKCHKRQKDVSCPIQVPTGKKQV 250
Db 1256 KKLFDLENINEETSVEVDRSFSSSKYNDVSVMIQIEDCNK----- 1297
QY 251 PLIPDLNQTGPNPFLSAVSNPEPSSHMVKSYSLLFR-----VTRPGRREFNG 301
Db 1298 -----NLNEK---NNKCRILLQNNIEMTMDIFVEYETESYRNTENEGNQCTDAGNTCNS 1350
```

```
QY 302 MING---ETNENIDVNEE---LPARRKRNREDGEKTFVAQMTVFQDKRRLQLLDGEYEV- 354
Db 1351 ESDGSDSSKNDIVYIHEBENGICPDIDQHNID-----LKLFSQFMKEGNTQIKEGLSDLT 1404
QY 355 ---AVQMEBECFISKRRATWETILDGKRLPPPTF--SQGPTLQTLQTLRWETNDKSTAP 409
Db 1405 CLEVKAETSTHVTMSNKQQLTANTQNIKDPDTFYLSTFQATASRNIRVSRSLNKAESL 1464
QY 410 IAKPLATR-----SESLHOENKPG-SVKPTQTIIVKESLTTDLOQRKEDTPNENRQKLR 464
Db 1465 LNKQWTEBELNFSLSLSELLPGIDIKKD---ISNHEVIENTERKKIYKESLIGTE 1521
QY 465 IFYQFLYNNNTQQTEARDDHLCPWCTLNCRLKLSLKLKLCRSF-----IFNY 515
Db 1522 NILLLIQORPESKIKKIKESAVLGFHTASGKKIETKESLDKVKNLFEKEQDNSEITNF 1581
QY 516 VYHPKGARIDVSNICYDG-SYAGNPQDIHQPGFAPSRNGVPVKETPITH-ILVCRP--- 570
Db 1582 SH--RGAKMSKDRBECKDQRELAOCGTTITTTTPEYEETHSSLEKCKLVSNELIALRPLL 1639
QY 571 ----KRT-----KASMSBFLSEDEGE---VEQORTYS-----SGHNR 600
Db 1640 SDNLYKQTEENLKSIDHASQKVDVHENTEKETAKKPTMTNQSTYSAIENSPLTFYTGHR 1699
QY 601 LYFHSDTCLPLRPQMEVDSDEKDPWMLREKTIITQIEFSD-----VNEGEKEVMKLV-- 654
Db 1700 KISVSEASL-----FEAKK---WLRE-----GSDDDQSRINAANKVCLKEYPD 1740
QY 655 -----NLHVMKHGFIADNOMNHACMLFVENYQGKIIKKNLCRN-FMLHLVSMHD---FN 704
Db 1741 DIVENPSCGSSNSALITENDKH-----LSEKQGSTYLSNSTMNSYSYHPGFCHSSEVYN 1796
QY 705 LISIMS-----IDKAVTKLREMQ-----OKEKGE 729
Db 1797 KSEYLSRSKIDNSGIEPVIKTRERKNIGFSEINSFGREADTDPQSVNEDICVEKLATNS 1856
QY 730 SASPAN-----EETEEQNGTANGFSEINSKE-----KALE--TDSV 764
Db 1857 SKCKNRKTAIKVAISDNNFNFTIQLKNSDNNNSVPAYSVNKRVFVAHQTKVTEGFTDNC 1916
QY 765 SGVSKQSKQK 775
Db 1917 SMVTKQNTKSK 1927

RESULT 12
Q8I525 PRELIMINARY; PRT; 5767 AA.
AC Q8I525;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hypothetical protein.
GN PFL1930W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Paine A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Bisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Partea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings I.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium falciparum.";
```



```
RL Nature 419:498-511 (2002).
DR EMBL: AEO14850; AAN36472.1; -.
KW Hypothetical protein.
SQ SEQUENCE 5767 AA; 691408 MW; E84A340441C67160 CRC64;

Query Match
Best Local Similarity 3.9%; Score 157.5; DB 5; Length 5767;
Matches 151; Conservative 151; Mismatches 268; Indels 249; Gaps 36;

QY 28 DLIEHEDNHDTPRVLEKQELQOPTY-----VALSYNRFMTDAAREQESLKK 79
Db 4056 DLINKEDLDIKNEMLKREELQDKIHIDELKNDMLKFKLYNAQNDIKVETKLUK 4115
QY 80 IQPKLS-----LTLSSSVSGNVSTPPRHSSGSLTPPVTPPTPSSFRSSTPTPTQIYR 135
Db 4116 IYEKLKDHNIIEFLDINNDRDNNISNKHSTNNI-----NSDIKQHT----- 4156
QY 136 FLTRNLIAPIFLHRTLTYMSHRNSTNIRK--TFKVDMLSKVE--KMKGEQSHSLSA 192
Db 4157 ---TNNI-----NNINNINNRLSLKRIIGLLGDEIGDENELDFHFDKENESLNL 4206
QY 193 HLQLTFTGFFHKNDKPSNSEQNSVTLEVLVKVCH---KKRKDVSCPIRQVPTGKKQ 249
Db 4207 D-SLEVTYTFNFKKIKLKNLNRKEDVIYKSYIEKKEIY-----KIE 4256
QY 250 VPLIPDLNOTKPGNFPPLAVSNEPEPSNSHMKYSYLLFRVTRPGRREFNGMNGETNE 309
Db 4257 LEKLNELLHVEQWNRKNDLELEKYKSEDTHTVKSLL-----RESEELLNEKNNK 4305
QY 310 NIDVNEELFARRKRNREDGEKTFVAQMTVPDKNRL-----QLLD-----GEYEV 354
Db 4306 IIELOOKLI-----ETSEYINIMI--DKNKLIKEKDYQKIDNLNLIINDYEK 4354
QY 355 AMQEMBECPISKRAWTETILDGKRLPPFPETFSQGTFLQTLRWGTGETNDKSTAPIAKPL 414
Db 4355 EIKEMNKELKLTAKSIEKMD-----DGEDIKKLKIDLE-- 4390
QY 415 ATNSESLSHQENKPSGVKPTQIANKESLTTLQTRKEKDTNENRQKLRIFQYQLYNN 474
Db 4391 -----AQELILDL-----KEK--NEELNNINI---ELINSN 4416
QY 475 TRQQTARDLDHCPWCTLCNCRKLYSLHLKLCRSHFIPVYVHPKGRDIVSINECYDG 534
Db 4417 ----XDMRLDIDVLLSNVDNLANKENILNEKKNDEQKN-----ELKIAYD- 4459
QY 535 SYAGNPQDTHROPGFAPSRNGPVKPTPITHILVCRPKRTKASMBFLES-EDGEVEQORT 593
Db 4460 -----QKVHE-----FNK-----LYNMLSIKWKKNKFDSEDMQNYTNDVNMOKS 4500
QY 594 YSSGHNRL--YFHSPTCLPLRQEMVDSEKDPMLREKTIITQIEFFSDVNEGEKEV 650
Db 4501 RNSINKYDINDFENDTIILNEKMIIESLNEKIAHY--EBENYKKEEL--VNKYKMI 4556
QY 651 MKLWN-LHYMK-----HGFADQNMNHACMLFVENYG---QKIICKNLCRN-FMLH 696
Db 4557 NELSNKIHIFEDNLDPLISNEFIKTNFTYQNLIDLNNSILYKRLKSLILVENQFYKY 4616
QY 697 LVSMHDF-----NLISMSIDKAVTKIREMOOK----- 724
Db 4617 IISIRDLWPEIITKSYSSNNNNNNNNIISNNSNTRESVLRMSLRKKGKFRKTASVTSNI 4676
QY 725 -----LEKGESASPAHEITEEQGTANGF--SEINSKE 756
Db 4677 YNDLFLSENVLSAEKENDIEMSFNKLNLNVIDDENSFE 4715

RESULT 13
Q9AHL0 PRELIMINARY; PRT; 1011 AA.
AC Q9AHL0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
```

```
DE LMP1.
GN LMP1.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_taxid=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3;
RA Dunn J.J.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF305603; AAK18795.1; -.
DR InterPro; IPR002713; PF.
DR InterPro; IPR006597; Sel_like.
DR Pfam; PF01846; PF; 4.
DR Pfam; PF00515; TPR; 9.
DR SMART; SM00671; SEL; 3.
DR SMART; SM00028; TPR; 8.
SQ SEQUENCE 1011 AA; 116004 MW; F6E8AD4365E5F38E CRC64;

Query Match
Best Local Similarity 3.8%; Score 157; DB 2; Length 1011;
Matches 160; Conservative 143; Mismatches 295; Indels 262; Gaps 43;

QY 25 TLADLIB-----HIEDNHDTPRVLEKQELQOPTYV-----ALSYNRFMTDAAREQE 74
Db 81 TLBIANSPPESKQLQDSANQILDKIKGQDNTKTNVNFDAFNRYIKDS----- 134
QY 75 SLKKKIQPKLSLTLSVSSRGNVSTPPRHSSGSLTPPVTPPTPSSFRSSTPTPTQIY 134
Db 135 -----TITENYSDRNDVIGIEDISEFKKSIPEKIKENTN-----PKEEQII 179
QY 135 RFLTRNLIAPIFLHRTLTYMSHRNSTNIRKTFKVDMLSKVEK-MKGEQSHSLSAH 193
Db 180 Q-----SPNPKLSVNDQKNLFN-----LEKLKKNLSGKNSSENILNN 216
QY 194 LQLTFTGFFHKNDKPSNSEQNSVTLEVLVKVCHK--RKDVSCPIRQVPT-GKKQV 250
Db 217 SQKI-----ENDQNTLSEKNS-----ENILKTPDMSKYNNNNNTLSKKIPSNOKES 267
QY 251 PLIPDLNQTGPNFPPLAVSNEPEPSNSHMKVS--YSLD-----FVTRPGRREFNGMI 303
Db 268 ELSF-----PSQTIIGKIYRP--YSYLIKKELEYLDDINTGRVTL-GKURLKELI 315
QY 304 -NGETNENIDVNEELPARRKRNREDGEKTFVAQMTVPDKNRLQLLDGEYEVAMQEMBE 362
Db 316 KKGSLNKFQKVNELI--ENSKNKE-----ASNLLTLIKKIDIEPNLINIPKD 360
QY 363 PISKRAWTETILDGKRLPPFPETFSQGTFLQTLRWGTGETNDKSTAPIAKPLATRNESL 422
Db 361 PY--KKEIFQDKEDKXPOYLEDL-----KSKVHSIPIDLENTKSR 400
QY 423 HQ-----ENKPGSVKPTQTIANK-----ESLTTDLOTRKEKDTPN--ENRQKLRI 465
Db 401 QQAIKDLNEFLKNNPNDQAASKTLAQAKIQLHLENLKSQVHSIKPIDLENTKSRQAID 460
QY 466 FYQPLYN-----NTRQOTE-ARDOLHC--PWCTLCNCRKLYSLHLKLCRHS 509
Db 461 LNEFLKNNPNDQAASKTLAQAKIQLHLENLKSQVHSIKPIDLENTKSRQAIDKL----- 516
QY 510 RFIENYVYHPKGRDIVSINECYDGSYAGN-----PDTHROPGFAPSRNGPVKRT 560
Db 517 EFLKN--NPNDAQASKTLAQAKIQLHLENLKSQVHSIKPIDLENTK-----SRQAIAK-- 567
QY 561 PITHILVCRPKRTKASMBFLESDEGEVEQORTYSSGHN-----RLYFHSPTCLPLRP 613
Db 568 -----DLNEFLKNNPNDQAASKTLAQAKIQLHLENLKSQVHS-----IKP 607
QY 614 QEMEVDSEKDPMLREKTIITQIEEPSDVNEGEKEVMLKWLNLHVMKHGFIADQNMNHAC 673
Db 608 ----IDLNTKYS-----RQQAIKDLNEFLKNNPNDQAASKT-----LAQAYENNGD 649
QY 674 MLFVENYQKIIKKNLCR-----NFWL-----HLVSMHDFNL-----I 706
```

```

Db 650 LKAEANAYEKIKLNTQEDHYKLGIIIRFKUKYEHSESFDQIKLDPKPKKALHNKGI 709
QY 707 SIMSDIKKAVTLKREMQ--KLEKGSASPAHEEITEQNGT-----ANGPSEINSKE 756
Db 710 ALMLNKNKAIESPEKAIQIDKNYSTAYYKGAIEKKNQMQQAFSEFKAYNLDKPEN 769
QY 757 KALETDSVS---GVSKQSKK 773
Db 770 YALKAGIVSNLGNPKQSEK 789

RESULT 14
OSIKPS
ID Q8IKP5 PRELIMINARY; PRT; 1163 AA.
AC Q8IKP5;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN PF14_0859.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,
RA Carlson J.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Shellen J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511 (2002).
DR EMBL: AE014825; AAN37172.1; -
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; zf_C2H2; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 1163 AA; 139588 MW; 6B37EB4621D0DAB1 CRC64;

Query Match
Best Local Similarity 3.8%; Score 157; DB 5; Length 1163;
Matches 170; Conservative 127; Mismatches 303; Indels 308; Gaps 47;

QY 19 CGLHFEPTLADLEHIEDNHIDTPRVLEKQBLQOPT-----YVALSYIN----- 62
Db 271 CGLRPMTSMKLNHLE-NHYKKSQFYLNNSSRSFTKSKKFLYLDHINLPIEVFVCKNY 329
QY 63 -----RFM-----TDAAREQESL-----KKKIQPK 83
Db 330 SIFEDFYNVVTKNWSNFQGSHEIEQILDNDNDNSYHNDDEKIYEEGEKKKQDESK 389
QY 84 LSLTLSSSVSRGNVSTPPRHSSGSLTPVTPPTPSGSFRSSTPTPTQIYRFLRNL 143
Db 390 NNDLSYSL-HTKIHTEHNKDN-----QESFLNN-----NTRKIQ 424
QY 144 APIFHLRLTYMHSRNSNTIK-----RKTFKVDMLSKVKMKGEBSHLSHLQTLF 198
Db 425 EKYNNYNNLNNNNNNNDVANLNFNFKKTYNID----- 459
QY 199 TGFHFKNDKPSNSENQSVTLVLLVKVCHKKKQVSCPIQVPTGKQVPLIPDLNQ 258
Db 460 -NFFH-----NKETIHMINPNAVDK--NKKKNI-----LND 490
QY 259 TKPGNFPPLAVSSNEFEPSNHHMVKSY-SLLPRVTRGRREFNGINGET-NENI----- 311

```

```

Db 491 REKG---SNEINFFDYIYGNENTYDVLYSNYYVTEDNSVLIN-YINGPTMYRNIMKCLE 546
QY 312 --DV-NEELPA---RRKRNREDGEKTFVAQMTVFQKNRLQLLDGEYEVAMQMBECCIS 365
Db 547 IKDIYVYKFPSPWPKESIN-----NFFIKRIEIDKD-----LIQ 583
QY 366 KKRATWETILDKRLLPFPPTFS-QGPTLQFTLRWTGETNDKSTAPIAKPLATRNSESL-- 422
Db 584 KKTNIFTNLSGR--TPSNTPFLPSYNIQDDLNNNNNNNNLKGKKNKIDIVINKNNLWN 641
QY 423 ---HQ-ENKPGSVKPTQTIIVKESLTTDLQ-----RKEKOTPN-----EN 459
Db 642 ICTTHEFYNSKGLCPINVFLSYKENTLHQHSDKNDHDDKESNNYPMQNSICTQES 701
QY 460 RQKLRIFYQFLYNNNTROQTEARDLHCPWCTLNCRLY-SLLKHLK-----LCHSRFI 512
Db 702 VENNHLVLSFL--NIFNQKYFAQTDIFKNILLFIRKHYFMILNELKNIDFNEULHSR-L 758
QY 513 FNYVHPKGARIDV-----SINECYDGSYAGNPODIHQPGCAFSPRNPVKRTPTIHL 566
Db 759 FQYIYNNYSKTISIFNNYKIDTCF-----LCKENFSFEYSVEYNDFYVTNVI 807
QY 567 VCRPKTKASMSEFLESEGEVQOQRTYSSG-----HRLYFHSHTCLPL 611
Db 808 CVDLK-----NVYENDDDVEDTEEDANKNYDIDMSIKRIDHMCDEYVYNNNSYDIDMKCLYE 863
QY 612 RPQEM-EVDSDEKDPMLREXTITQIBEF-----SVNVE--GEKVMKLNHLNLMKH- 661
Db 864 KHKELNELLYENKE---MSDDAITQVELLNNKNNYNEINECSNHNKNNINITYDNI 920
QY 662 -----GFIADQNMMHACMLFVNYGQKIKNKCRNFMHLVSN-----H 701
Db 921 NINNIFTDIKKIIDDN-----CFL-----GNTIIESNMDCNFMHNIINKKCKKDHIMNH 970
QY 702 DFNLSIMSIDKAVTKLRMOCKLEKSGSASPANEITEEQNGTANGSEINSKEKALET 761
Db 971 YNNLFKNLCIPYDTL---DILHRIKKG-----DITGKNTMTNNTNNTNDEGYTN 1019
QY 762 DSVSGVSK 769
Db 1020 EEDNDFLK 1027

RESULT 15
Q9W4G6
ID Q9W4G6 PRELIMINARY; PRT; 1893 AA.
AC Q9W4G6;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Cg2861 protein.
GN Cg2861.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Franchoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:46:43 ; Search time 45 Seconds
(without alignments)
1659.769 Million cell updates/sec

Title: US-09-874-162A-8
Perfect score: 4078
Sequence: 1 MTGIAAASPFNTCRFGGCG.....KALETDSVSGVSKQKQKL 776

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*

1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 176.5 | 4.3 | 1142 | 2 S59359 | GIN4 protein - yea |
| 2 | 166 | 4.1 | 1507 | 2 T42631 | breast cancer tumo |
| 3 | 163 | 4.0 | 1314 | 2 T09481 | mating type silenc |
| 4 | 155.5 | 3.8 | 1755 | 2 S57045 | TyB protein - yea |
| 5 | 154 | 3.8 | 1271 | 2 A45555 | glutamate rich pro |
| 6 | 152 | 3.7 | 1119 | 2 B70126 | surface-located me |
| 7 | 152 | 3.7 | 1749 | 2 S69972 | TyB protein - yea |
| 8 | 152 | 3.7 | 1957 | 2 T38077 | hypothetical coile |
| 9 | 150 | 3.7 | 912 | 2 B44450 | ubiquitin-specific |
| 10 | 149 | 3.7 | 1145 | 2 T18235 | transcription acti |
| 11 | 148 | 3.6 | 863 | 2 T51002 | hypothetical prote |
| 12 | 147.5 | 3.6 | 1368 | 2 T18371 | probable glutamate |
| 13 | 147.5 | 3.6 | 1755 | 2 S63980 | TyB protein - yea |
| 14 | 147.5 | 3.6 | 1790 | 2 S67593 | transport protein |
| 15 | 146.5 | 3.6 | 1235 | 1 PWSYH | potassium transpor |
| 16 | 145 | 3.6 | 1744 | 2 JH0720 | tanabin - African |
| 17 | 145 | 3.6 | 1953 | 2 S63244 | BN11 protein - yea |
| 18 | 143.5 | 3.5 | 698 | 2 S45027 | hypothetical prote |
| 19 | 143 | 3.5 | 1640 | 2 A24594 | probable major sur |
| 20 | 142 | 3.5 | 1278 | 2 B24671 | TyB protein - yea |
| 21 | 142 | 3.5 | 1738 | 2 T14466 | interaptin - slime |
| 22 | 141.5 | 3.5 | 1263 | 2 T15496 | hypothetical prote |
| 23 | 141.5 | 3.5 | 1328 | 2 S52481 | TyB protein - yea |
| 24 | 141.5 | 3.5 | 1755 | 2 S50641 | TyB protein YER138 |
| 25 | 141.5 | 3.5 | 1755 | 2 S63975 | TyB protein - yea |
| 26 | 141 | 3.5 | 1827 | 2 T18270 | hypothetical prote |
| 27 | 141 | 3.5 | 1875 | 2 S38173 | myosin-like protei |
| 28 | 140.5 | 3.4 | 4688 | 2 F82885 | hypothetical prote |
| 29 | 140 | 3.4 | 1631 | 1 SAZQK1 | major merozoite su |

30 139.5 3.4 1085 2 S62516 hypothetical coile
31 139.5 3.4 1755 2 S69949 TyB protein - yea
32 139.5 3.4 2500 1 MHUE2 HIV-SP2 enhancer-b
33 138.5 3.4 582 2 E84771 fertilization-inde
34 138.5 3.4 1755 2 S50663 TyB protein - yea
35 138.5 3.4 1793 2 S52601 TyB protein - yea
36 137.5 3.4 1196 2 S40908 TyB protein - yea
37 137.5 3.4 1755 2 S69839 TyB protein - yea
38 137.5 3.4 1755 2 S69960 TyB protein - yea
39 137.5 3.4 1755 2 S69979 TyB protein - yea
40 137.5 3.4 1755 2 S69955 TyB protein - yea
41 137.5 3.4 1756 2 S69983 TyB protein - yea
42 136.5 3.4 2954 2 T14156 kinesin-related pr
43 136.5 3.3 791 2 T20815 hypothetical prote
44 136.5 3.3 1755 2 S70238 TyB protein - yea
45 135.5 3.3 1295 2 T24587 hypothetical prote

ALIGNMENTS

RESULT 1

S59359
GIN4 protein - yeast (Saccharomyces cerevisiae)
N/Alternate names: protein YDR507c
C/Species: Saccharomyces cerevisiae
C/Date: 30-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 24-Sep-1999
C/Accession: S59359; S69565
R/Longtine, M.S.; Pringle, J.R.
submitted to the EMBL Data Library, August 1995
A/Reference number: S59359
A/Accession: S59359
A/Molecule type: DNA
A/Residues: 1-1142 <LON>
A/Cross-references: EMBL:U33140; NID:g992650; PIDN:AAA75513.1; PID:g992651
R/Dietrich, F.S.
submitted to the EMBL Data Library, August 1995
A/Description: The sequence of S. cerevisiae cosmids 8166, 9787, 9717, and lambda 3073.
A/Reference number: S69553
A/Accession: S69565
A/Molecule type: DNA
A/Residues: 1-1142 <DIE>
A/Cross-references: EMBL:U33057; NID:g927764; PIDN:AAB64949.1; PID:g927777; MIPS:YDR507c
C/Genetics:
A/Gene: SGD:GIN4
A/Cross-references: SGD:S0002915; MIPS:YDR507c
A/Map position: 4R
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C/Keywords: ATP
F/17-289/Domain: protein kinase homology <KIN>
F/25-33/Region: protein kinase ATP-binding motif

Query Match 4.3%; Score 176.5; DB 2; Length 1142;
Best Local Similarity 20.8%; Pred. No. 0.0037;
Matches 152; Conservative 102; Mismatches 264; Indels 213; Gaps 35;
QY 41 DPRVLEKQEIQQP-----TVVALSYINRFMTDARREQSLKKKIQPKLSLTLS--SVS 93
DB 337 DPEGI-KEKLREPAGNAEKLTYALLY--RPFKCDT---OKELIKQQQVKRQSISSVSVS 389
QY 94 RG-NVSTPP-----RHSSGSLTTPVTPPI-----TPSSSFRSSTPTPTQIVRFLRTR 140
DB 390 PSKVVSTTPQRRNRRESLISVTSRKPPIFNKFTASSASSNLTTPGSSKRISKPNSSK 449
QY 141 NLIAPIFLHRTLTVMYSHRNSTNI--KRKTFKVDMLSKVKMKGEQES----- 187
DB 450 KKLSTIVNQSSPTPAS--RNKRASVINVEKQKQASIPSTTKNKRSSRSIKRMSLIPSMK 508
QY 188 -HSLSAHLQIQTFTGFFHKNDKPSNSEQNSVLEVLVVKV-----CHKKRKDVSC 238
DB 509 RESVTTKLMSTYAKLAEDDDWEYIERETKRTSSNFATLIDEIFEYKYEKQIRKEKELE 568
QY 239 PIQVPTG-----KKQVPL-IPDLNQ----- 258

```

Db      569 KVRKAREELERRRRKQEKERARKLLEKEDLKKQBELKKQIEIDISLDEQLSKHKE 628
QY      259 -TKPQNPFLAVSSNEFPFNSHMVKSYSLLFR-----VTRPGRREFNGMINGETNENI 311
Db      629 EKLDGNIRISAPMENEENKINHLVDINILRRRNFSLQTRPVSRDLDPGIMFSPTEEV 688
QY      312 DVNEELPARRKRNREDGKTFVQAM-----TVFDKNRRLQLLDGEVE--VANQEM-EE 361
Db      689 SPVE--PKRTENERUTTKKILLETIRRSKFLGSSFNIDKELKMKMEYPSIIAQRLSEE 746
QY      362 CPISKRRATWETIL--DGKRLPPPE--TFSGQPTLOFTLRWTGETNDKSTAPIAKPLAT 416
Db      747 RVVSDNDGYESLILPKDNGVSQLKSTATTAPVSDGRLRKISE-----IRVQFT 798
QY      417 RNSBSLHQENKPGSVKPTQTAVKESLTTDLQTRKEKDTFNNRQKRLIFYQFLYNNNR 476
Db      799 RKSRSFSESNKLSV--LSMYSTKESFTNLVDILKNGLDVNNQSQRI-----845
QY      477 QOTEARDLHCPWCTLNCRKLYSLKLKHLKCHSRFIFNYVYHPKRGARIDVINECYDGSY 536
Db      846 PTPRSADD-----SEFLFE-----TVNE--EABY 867
QY      537 AGNPQDIHR-----OPGFA--FSRNGPVKRTPIITHILVCRPKRYKAMSE 579
Db      868 TGNSSNDERLYDVGDSTIKDKSALKLNFAFRNGSNEAKQTDNLHLPLILPLINGDNLK 927
QY      580 FLESDEGEVQORTYS---SGHNLVPHSDTCLPLRQEMVEVDSDEKPEW-----628
Db      928 -QNSGEQOAHPIKSMIPESGSS---HTE-----KEENEKEKEKPEQHQBEDQ 976
QY      629 -LREKTIQIE 638
Db      977 EKREKVDDME 987

RESULT 2
T42631
breast cancer tumor suppressor BRCA2 - dog (fragment)
C/Species: Canis lupus familiaris (dog)
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 24-Nov-2003
C/Accession: T42631
R/Bignell, G.R.; Micklem, G.; Stratton, M.R.; Ashworth, A.; Wooster, R.
submitted to the EMBL Data Library, July 1996
A/Description: BRCA2 contains a novel repeat that is conserved in other mammian species
A/Reference number: Z22184
A/Accession: T42631
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-1507 <BIG>
A/Cross-references: EMBL:Z75664; NID:e996328; PID:e251256; PIDN:CAA99994.1
C/Genetics:
C/Suprafamily: breast/ovarian cancer tumor suppressor protein, BRCA2 type
C/Keywords: tumor suppressor

Query Match      4.1%; Score 166; DB 2; Length 1507;
Best Local Similarity 18.9%; Pred. No. 0.025;
Matches 180; Conservative 143; Mismatches 384; Indels 244; Gaps 43;

QY      23 FPTLADLIEHDNHDIDPRVLEKQELQQPTTYA--LSYINRFMTDAARQESLKKKIQ 81
Db      373 FKTASNKIKLSENNIRKSKMLFKDIBEHYPTNLACLEIVTSSLESQKKPSKS--HALD 430
QY      82 PKLSITLSSVSRGNVTPPHRSGSLTPPTPTTPSSSPRSSTPTPEPTOI-----133
Db      431 POSINIIISGFVQ--NSTVVSDESHTAPPTLSLKQDPDSNRNLTPSQKABITELSTILE 488
QY      134 -----YRLRTNLNLIAPFLHRTLTMYGHRNSRT--NIKRTFFKVDMM-----LSKV 178
Db      489 ESGSFQFTQFRK---PSHIQKPPFEMPENQLITLSTSEKWKDDDLHLTNAPSISQV 545
QY      179 ERMK-----GEQSHSLISA-----HLQITFTGFFH-KNDKPSFNSE-----213

```

```

Db      546 DSKKSEGIIGKQKQACFLSRTSCNRSASGYSTDKNEVEFRGYSGTGTLNVGSEALQKA 605
QY      214 -----NQNSVTLEVLVVKVCHKKQDVSCPIQVFTGKQVPLIPDLNQTQKPN 263
Db      606 KKLFSLENIENETSVEVD-----RSFSSSKYNDVSMIQIEDCNCKNLEPN 654
QY      264 FPSLAYSSNEFSPNSHMVKSYSLLFR-----VTRPGSRETFNGMING---ETNENI 311
Db      655 KRL--ILQNIEMTTDIFVEETESYRRNTENEGNOCTDAGTNCNSEDGSDSKNTIV 713
QY      312 DVNEE--LPARRKRNREDGKTFVQAMTVFQKRLQLLDGEYEV-----ANQEMEECP 364
Db      714 YIHEENGPLCIDQHNID-----LKLFSQFMKEGNTQIKEGLSDLTCLVNMKABETSHV 767
QY      365 SKKRAWTETILQKRLPPETP--SQGPTLOFTLRWTGETNDKSTAPIAKPLATN----418
Db      768 TWSNKOQLTANTGONIKDFTFYLSQFASRNKIRVSRESLNKARSLNKKQWTEBELANF 827
QY      419 SSSLHQENKPG-SVKPTQTIIVKESLTTDLQTRKEKDTFNNRQKRLIFYQFLYNNNRQ 477
Db      828 SDSLSNELLPGIDIKTD--ISNHEVIENTERKDKITKESDLIGTENILLILQORPSSK 884
QY      478 QTEARDDLHCPWCTLNCRKLYSLKLKHLKCHSRF-----IFNYVYHPKRGARIDVSI 528
Db      885 IKKIKESAVLGPHTASGKKIETIKESLQKVLFEKQDNSEITNFSH--RGAKMSKDR 942
QY      529 NECYDG-SYAGNPQDIHQPGFAFRNGVVKRTPIITH-ILVCRP-----KRT-----573
Db      943 ECKQORELACGTTEITTYPEETHSSLEKKLVNBIALPRLSLDNLKQYENLKI 1002
QY      574 -----KASMSFLESDGE---VEQRTYSSGHRN-LYFHSDTCLPLRQEMVEVDS 622
Db      1003 SDHASQKVDVHENTKETAKKPTMTYNQSTYSIAENSPLTFTQDT-----EEKFSVSEA 1056
QY      623 E--KQPEWLREKTIQIEFSD-----VNEGEKEVMKLM-----NLHVMKHGFIADNQ 668
Db      1057 SUFEAKKWLRE-----GEWDDQSERINNAKVNCLKEYPDDYVENPSCGNSNSAITEND 1110
QY      669 MNHACMLFVENYQKIIKKNLCRN-FMLHLVSMHD---FNLISIMS-----IDKAVT 716
Db      1111 KXH---LSEKQGSTVLSNSTMSNSYVHPGCHSSEVNVKSEVLSRSKIDNSGIEPVIK 1166
QY      717 KLREMQ-----QKLEKESASPAN-----735
Db      1167 NIREKNIGFSIMSGREADTDPQSVNEDICVEKLATNSSCNKNNTAKVAISNNFN 1226
QY      736 --EETEEQNGTANGFSEINSKE-----KALE--TDSVSGVSKQSKQK 775
Db      1227 TIQKLSNDSNNSVPAYSTVNSKRVFAHQTKVTEGTDNCMSVTKQNTKSK 1277

RESULT 3
T09481
mating type silencing regulator SIR4 - yeast (Kluyveromyces marxianus var. lactis)
N/Alternate names: silent information regulator 4
C/Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C/Accession: T09481
R/Astrom, S.U.; Rine, J.
Genetics 148, 1021-1029, 1998
A/Title: Theme and variation among silencing proteins in Saccharomyces cerevisiae and Kl
A/Reference number: Z16686; MUID:98198828; PMID:9539421
A/Accession: T09481
A/Status: translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-1314 <AST>
A/Cross-references: EMBL:AF035007; NID:g3002787; PID:g3002789
A/Note: the source is designated as Kluyveromyces lactis
C/Genetics:
A/Gene: SIR4
C/Function:
A/Description: involved in both silencing and telomere length maintenance

```

C;Keywords: transcription regulation

Query Match 4.0%; Score 163; DB 2; Length 1314;
Best Local Similarity 19.9%; Pred. No. 0.032;
Matches 169; Conservative 140; Mismatches 338; Indels 204; Gaps 39;

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|-----|-----|---|---|---|---|---|---|---|---|-----|-----|---|-----|---|---|---|---|---|---|---|-----|---|---|---|-----|---|---|---|---|---|---|---|---|-----|---|-----|---|-----|---|---|---|-----|---|---|---|---|---|---|---|-----|-----|---|---|-----|---|---|---|
| QY | 43 | RVL | E | K | B | E | L | Q | B | E | L | --- | Q | O | P | T | I | V | A | L | S | V | I | N | E | M | T | D | A | A | R | R | E | S | L | K | --- | K | I | O | P | K | L | S | L | T | L | S | S | V | S | --- | R | G | 95 | | | |
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | | | | |
| Db | 181 | R | V | E | D | E | V | L | R | Q | O | S | S | L | S | S | G | R | I | K | H | I | E | S | P | F | O | R | N | S | K | P | V | S | I | V | K | A | P | N | K | V | N | F | A | F | E | S | I | D | L | 240 | | | | | | |
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | | | | |
| QY | 96 | N | V | S | T | P | P | R | H | S | G | S | L | T | P | P | T | P | S | S | R | S | T | E | P | T | Q | I | Y | R | F | L | R | N | --- | L | I | A | P | I | F | L | 284 | | | | | | | | | | | | | | | |
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | | | | |
| Db | 241 | N | L | I | T | R | A | S | S | G | --- | T | P | S | K | S | A | V | P | A | P | --- | K | T | K | T | D | S | O | V | S | K | I | A | N | E | I | O | 284 | | | | | | | | | | | | | | | | | | | |
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | | | | |
| QY | 149 | H | R | T | L | T | Y | M | H | R | N | S | R | N | I | K | R | T | F | K | V | D | M | L | S | K | V | E | K | O | E | S | H | S | L | A | H | L | O | T | F | T | G | P | F | H | K | D | K | P | 208 | | | | | | | |
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | | | | |
| Db | 285 | N | S | N | K | S | R | I | S | K | A | N | S | --- | P | I | O | R | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T | T |
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | | | | |
| QY | 209 | S | P | N | S | E | N | E | Q | N | S | T | L | E | V | L | L | A | V | K | C | H | K | R | D | V | S | C | P | I | O | V | P | T | G | K | O | V | P | L | I | P | D | L | N | Q | T | K | P | N | F | S | L | A | 268 | | | |
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | | | | |
| Db | 338 | S | E | N | N | D | T | G | A | V | K | S | S | H | V | E | I | G | S | K | R | D | T | H | T | S | F | M | E | L | P | S | M | Q | A | T | I | A | L | K | T | S | L | V | K | N | S | A | Q | I | E | 397 | | | | | | |
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | | | | |
| QY | 269 | V | S | S | E | F | E | P | S | N | H | M | V | K | S | L | L | F | R | T | P | R | R | E | F | N</ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4

REPORT
S57045

TyB protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty1.JR

N;Alternate names: protein J1560; protein YJR027w

C/Species: *Saccharomyces cerevisiae*

C;Date: 08-Jul-1995 #sequence revision 08-Sep-1995 #text change 17-Mar-1999

C:Accession: S57045; S60511
R:Zagalski, M.; Babinska, B.
Submitted to the Protein Seq
A:Reference number: S57040
A:Accession: S57045
A:Molecule type: DNA
A:Residues: 1-1755 <HUA>
A:Cross-references: EMBLJ24
A:Note: biosynthesis of thi
R:Zagalski, M.; Babinska, B.
Yeast 11, 1179-1186, 1995
A>Title: The sequence of 24
A:Reference number: S60503;
A:Accession: S60511
A>Status: nucleic acid sequ
A:Molecule type: DNA
A:Residues: 'TGTGTT', 435-1
A:Cross-references: EMBLX8
A:Note: the nucleotide sequ
A:Note: the difference at t
C:Genetics:
A:Map position: 10R
A:Mobile element: retrotran
C:Superfamily: TvB protein

| | | | | |
|-----------------------|--------------|-----------------|-----------------|--------------|
| Query Match | 3.8% | Score 155.5; | DB 2; | Length 1755; |
| Best Local Similarity | 22.6%; | Pred. No. 0.14; | | |
| Matches 159; | Conservative | 99; | Mismatches 232; | Indels 213; |
| | | | | Gaps 44; |

| | | | | | | | |
|----|------|--|----------------------------------|------------------------|---------------------|-------------------|------|
| Qy | 38 | ITDTPRVL---- | EKQELQOPTVVALSY---- | INRFMTD----- | AARREGESLUKKI----- | 80 | |
| Db | 881 | VDTNYVILQKESRLQDFNYDALTDFEDLNRLTASYQSF | AGNEQESLNDLINESDHDF | 940 | | | |
| Qy | 81 | ----- | QPKLSLTLSSSVSRGNVSTPP----- | RHSSGSLTTP-- | VTPEITPS-- | 119 | |
| Db | 941 | QSDIELHPEQPR-- | NVLSKAVSPTD-- | STPSTHTEDSKRVS | KTNIRAPREVDPMISSENI | 997 | |
| Qy | 120 | -- | SPFSSTSTPTQYRFLRTEN----- | LIAPFLHRLITYMSEN | SRTNIRKKT | 168 | |
| Db | 998 | LPKSKRSSTP-- | QISNIBSTGGMHKUNVLLAPMSQSNT-- | ----- | HESSHAS-KSKD | 1047 | |
| Qy | 169 | FKYDDMLSKYERKMGQESHLSAHLQLTGTGFHKNDKSP | NSENSEQNSVTLVLLVKY | 228 | | | |
| Db | 1048 | FRHSDSYS----- | ENETNITNVPISSTG-- | GTNNKTVPQISDQETE----- | KRI | 10899 | |
| Qy | 229 | CHKRKB-DVS----- | CPHQVPTGKKQ----- | ----- | VPLIPDLNQTKGNFPFSLA | 268 | |
| Db | 1090 | IHKSPSIDASPPENNASHNIVPIKTPTVSEQNTBES | IIADLPL-- | POLPESPTPEFPD-- | 1146 | | |
| Qy | 269 | VSNFEPSPNSHMVKYSLLFVRVTRPGRRFNGINGET | NIDNIVNSEL | PARRKRNBEHG | 328 | | |
| Db | 1147 | PPKXELPPINSHOTNS----- | ----- | SLGGI-- | GDSNAYTTINS----- | KKRSLEDN | 1185 |
| Qy | 329 | EKTF-VAQMTFVDFKNRRLQLLDGEYEVAMQEMEB | CPISKRR----- | ATWETILDGKRLPPF | 383 | | |
| Db | 1186 | ETILKVSIRDWTNTKNMR----- | ----- | SLEPPRSKKRIHLIAAVKAV-- | KSIKPI | 1228 | |
| Qy | 384 | ETPSQPTLOF-- | TLRWGTGNTDKSTAPIAKPIATRNSES | HLQBNKPGSVKPTQTI | AVKE | 441 | |
| Db | 1229 | RT----- | TLRYDEAITYNKDIKE----- | ----- | KYIEAYHKE----- | VNQLLKMKT | 1266 |
| Qy | 442 | SLATDLQTRKEKDTPENROKLAIFYOFLYNNNT-- | ----- | ROQTEARDDLHCP----- | WCTL | 492 | |
| Db | 1267 | WDTDEYDKREID----- | PKRVINSMFIENKKRDGTHKARFVARGDI | QHPDVTYDTGMQS | 1320 | | |
| Qy | 493 | NCKRLSYLLKHLKLCHSRFFITNVIYHPKGARIDV | SINECYDGSYAGNPQDIHQPGFAPS | 552 | | | |
| Db | 1321 | NTVHHVALMTSLSLDNL----- | NYII----- | TOLDIS-- | SAY-- | YADIKESILIRPPPHLG | 1369 |
| Qy | 553 | RNGFVKRTP:ITHILVCRPKRTKASMGFELES----- | EDGEVEQBOQRTYSSGHNRLFYHSDTCL | 609 | | | |
| Db | 1370 | MNDKLLRLKSKSHYGL----- | KOSGANWYETIKSYULIKOCGMBEVAGWSC-- | VFANSOVTI | 1432 | | |

```
QY 610 PLRQEMVEYDSEKDPWLRKTIQIEFSD-----VNEGEKE 649
      : : : : : : : : : : : : : : : : : : : : : :
Db 1423 CLFVDDMLFSKDLN-----ANKLIITLKKQYDTKIINLGS 1461

RESULT 5
A4555
glutamate rich protein - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C:Accession: A45555; 527831
R:Porre, M.B.; Dziegiel, M.; Hogn, B.; Petersen, E.; Rieneck, K.; Riley, J.F.;
J.
Mol. Biochem. Parasitol. 49, 119-131, 1991
A:Title: Primary structure and localization of a conserved immunogenic Plasmodium falciparum life cycle.
A:Reference number: A45555; MUID:92131041; PMID:1775153
A:Accession: A45555
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1271 <BOR>
A:Cross-references: EMBL:M59706; NID:G160311; PID:G160312
A:Note: sequence extracted from NCBI backbone (NCBI:77801, NCBI:77802)

Query Match 3.8%; Score 154; DB 2; Length 1271;
Best Local Similarity 18.0%; Pred. No. 0.11;
Matches 150; Conservative 123; Mismatches 270; Indels 290; Gaps 40;

QY 31 EHIENHIDTDPRVLE-----KQELQOPTYVALSYINRFTMDAARREQESLKKKIQ 81
      : : : : : : : : : : : : : : : : : : : : : :
Db 258 EHLKDEKLEPLVHENLSIPNDPIEQILNQ-----EQET---NIQ 296

QY 82 PKLSLTLSSVSRGNVSTPPRHSSGSLTPPVTPITPSSSRSTPTPTPTQIYRFLTRN 141
      : : : : : : : : : : : : : : : : : : : : : :
Db 297 EOLYNEKONVEEKQNSQIP-----SLDLKEPTNEDILP----- 330

QY 142 LIAPFLHRTLYMHRNRTN-IRKTKFKVDMLSKVEKMG--EQBSHLSLAHLQLTF 198
      : : : : : : : : : : : : : : : : : : : : : :
Db 331 -----HNPLENIQSESEINHVDHALPKNIDIKLDNKEHLDQSQHNINVLQENN 382

QY 199 TGFPHNDKPPNSN-SONSVTLVLVVKVCHKKRQDVSCPIRQVPTGKQVPLIPDLN 257
      : : : : : : : : : : : : : : : : : : : : : :
Db 383 INNHOLEPQEKPNIESFPKNIDSIIL-----PENVETEBIIDVP 424

QY 258 QTKPGNFFSLAVSNFE-----PSNSH-MYKSYSLFLVTRPGRREENGMINGETNEN 310
      : : : : : : : : : : : : : : : : : : : : : :
Db 425 SPKSHNHTEETSESEHEEAVSEKNAHETVEHEETVQSENPKEADNDGNSQNGNE 484

QY 311 IDVNEELPARRKRNREDGEKTFVAQMTVDFNRRLLQLDGEYE--VAMQEMEECPISKRR 368
      : : : : : : : : : : : : : : : : : : : : : :
Db 485 LNEENFVES-EKSEHEAAE-----NEESSLEEGHHEIVPEONNEESGESK-- 529

QY 369 ATWEIILD-----GKPLPFETFSQPTQLQTLRWGTETN-----DKS-TAPI----- 410
      : : : : : : : : : : : : : : : : : : : : : :
Db 530 -----LVNDDEGGFEEAHNFSSVS-----NSELNENFVESQKSVTEPAHEBEV 577

QY 411 ---AKPLATRNSESLSHQNKPGSVKPTQTIAGKSLTLDLQTRKEK-----TPNE 458
      : : : : : : : : : : : : : : : : : : : : : :
Db 578 SEESNPEAENEESEIEAHQEEIVPEON--DBESGESGLVDNEBGPPEEPNHEEFPDQ 635

QY 459 NRQKURIFYQFLYNNNTQQTEARDLHCPWCTLNCRKLYSLKHLKLCRFRFTFNTVYH 518
      : : : : : : : : : : : : : : : : : : : : : :
Db 636 NDSEL-----SENELVESE-----KSVSEPAHEIVEIVSEKVSSEPAEH 673

QY 519 PKGARIDVINECYGSVAGNPQDTHROPGFAS--RNGPVKRTPIITHI--LVCRPKRTK 574
      : : : : : : : : : : : : : : : : : : : : : :
Db 674 -----VEI-VSEKTSPEAHEVESVSESNNEPEKKGDPVSPFEEIEKVDVQPKIVD 727

QY 575 ASMSE--FLESBDGVEQOQTYSSGHNRLYFTHSDTCLPLRQEMEVD-----SEDBK----- 625
      : : : : : : : : : : : : : : : : : : : : : :
Db 728 LQIIEPNFVDSQNPQE-----PVEPSFVKIEKVPSEENKHAUSD 767

QY 626 PEWLREKTIITQIEEFS-----VNEGEKEMKLNHLVHKHGFADNQNHACML 675
```

```
Db 768 PEVKEKENVSEVVEEKQSQESVEEIPVNEDEBFDVHTEQ-----DLDH----- 812
      : : : : : : : : : : : : : : : : : : : : : :
QY 676 FVENYQKIIKKNLCRNFMLHLVSMHDFNLISIMSDKAVTKLRMOOKLEKGESASPAN 735
      : : : : : : : : : : : : : : : : : : : : : :
Db 813 -----KTVDFEIVEVEEIPSELHENEVAHFEI 839

QY 736 ERITE-----EQNGTANGFSEINSKEA-----LETDSVSGVSKQSKOKL 776
      : : : : : : : : : : : : : : : : : : : : : :
Db 840 VEIEEVFPENQN---NEFQEIENEDDKSAHQIEHVEVEEI--LPEDDKNEKV 887

RESULT 6
B70126
surface-located membrane protein 1 (lmp1) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 11-Jan-2000
C:Accession: B70126
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Yugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: B70126
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1119 <KLF>
A:Cross-references: GB:AE001131; GB:AE000783; NID:G2688098; PID:AAAC66595.1; PID:G268810
A:Experimental source: strain B31
C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat
F;742-774/Domain: tetratricopeptide repeat homology #status atypical <TT1>
F;775-808/Domain: tetratricopeptide repeat homology <TT2>
F;809-842/Domain: tetratricopeptide repeat homology <TT3>
F;843-876/Domain: tetratricopeptide repeat homology <TT4>
F;911-943/Domain: tetratricopeptide repeat homology <TT5>
F;944-977/Domain: tetratricopeptide repeat homology <TT6>
F;979-1010/Domain: tetratricopeptide repeat homology #status atypical <TT7>
F;1011-1044/Domain: tetratricopeptide repeat homology <TT8>

Query Match 3.7%; Score 152; DB 2; Length 1119;
Best Local Similarity 18.1%; Pred. No. 0.13;
Matches 147; Conservative 131; Mismatches 312; Indels 220; Gaps 33;

QY 74 ESIKKKI-----QPKLSLTSSVSRGNVSTPPRHSSGSLTPPVTPPI 116
      : : : : : : : : : : : : : : : : : : : : : :
Db 198 EKLKQNLGSKNSNINLDSQKIENDKQNTNLKSKNSNENILKTPDNKYSNNNTSLK 257

QY 117 TPSSSRSTPTPTQIYRFLTRNLIAPIFLHRTLYMHSNRTNIRKTRFKVDDMLS 176
      : : : : : : : : : : : : : : : : : : : : : :
Db 258 KISSNQKESLSPP-----SQTIIGKIY--RPVSYL-----IKKELYILDDIN 300

QY 177 KVEKMGGESESLSAHLQLTFTGFHKNDKXSPSNEQNSVTLVLVVKVCHKKRQDV 236
      : : : : : : : : : : : : : : : : : : : : : :
Db 301 TCRVTILGKVRLK-----ELIKKLSNPKQVKVNEIENSKNEASNLTLTL-----KKDI 350

QY 237 SCPIROVPTG--KKQVPLIPDLNQTQKPGNFPSPSLAVSNSEFEPNSGHMVKSYSLLFRVTRP 294
      : : : : : : : : : : : : : : : : : : : : : :
Db 351 EPNLINPKDPYKKEIFQL-DKEDKKPQVLEDLKSQVHSIKPIDLENTKSRQAIAK---- 405

QY 295 GRREFNGMNGETNENIDVNEELPARRKRNREDGEKTFVAQMTVDFNRRLLQLDGEVEV 354
      : : : : : : : : : : : : : : : : : : : : : :
Db 406 -----DLNEEL-----KNNPNDAQ-----ASKLAQANKIQHLEDLKSQV 440

QY 355 AMQEMEECPISKRAETWETILDGKELPPF-----ETFSQPTQLQTLRWGTETN 403
      : : : : : : : : : : : : : : : : : : : : : :
Db 441 ---HSIKPIDLENTKSRQAIAKD--LNEFLKNNPNDAQASKLAQANKIQHL-----ED 489

QY 404 DKSTAPIAKPLATRNSESLSHQ-----ENKPGSVKPTQTIYK-----ESLTDLQ 448
      : : : : : : : : : : : : : : : : : : : : : :
Db 490 LKSKVHSIKPIDLENTKSRQAIAKDINEFLKNNPNDAQASKLAQANKIQHLEDLKSQV 549
```



```

449 QY TRKEKPTFN--ENRQKLRIRYQFIYNNNTTQQTEARDLHCPWCTLNCRKLYSLLHLKL 506
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
550 Db SIKPIDLNTKSRQQAIKDLNEFLKNPNDAQASK-----TLAQANKI 592
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507 QY CHSRFIENVYVHPGASIDVSINCYDGSYAGNPQDTHRPGFAFSRNGPVKRTPIPHIL 566
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
593 Db QHLEDLSKQVHSIK--PIDLNTKSRQQAIKDLNEFLKNPNDAQASKTLAQANKIQHLE 650
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 QY VCRPKF-----RTKA-----SMSEFLESDGEVEQORTSYSSGHN-----RLYF 603
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
651 Db DLKSKVHSIKPIDLNTKSRQQAIKDLNEFLKNPNDAQASKTLAQANKIQHLEDLSKV 710
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
604 QY HSDTCLPLRQEMEVDSDEKDPWLBREKTIQTIERESDVNVEGEKFMKLNHLVWKHGF 663
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
711 Db -----IKP-----IDLNTKS-----RQAIAKDLNEFLKNPNDAQASKT----- 747
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
664 QY IADQNMHACMLFVENYQGIKKNNLCR-----NFML-----HLVSMHDFNL----- 705
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
748 Db LAQAYENNGLDLKAENAYEKILKLTNQDEHYKLGIRFKLKKVHSIESFDQTIKLDPK 807
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
706 QY -----ISIMSIDKAVTKLRMQQ--KLEKGESASPANEIITEFQNGT-----A 746
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
808 Db HKKALHNKGIALMNLNKKKAIESFEKAIQIDKNYGTAYVQKGAIEKNGDMQQAFAFSPK 867
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
747 QY NGFSEINSKEKALETDSVS---GVSKOSKK 773
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
868 Db NAYNLDKNPYALKAGIVSNNLGNFKROSEE 897
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

RESULT 7
S69972
TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.N
N/Alternate names: protein N2453
C/Species: Saccharomyces cerevisiae
C/Date: 09-Mar-1996 #sequence_revision 06-Sep-1996 #text_change 20-Jun-2000
C/Accession: S69972
R: Berger, P.; Doignon, F.; Crouzet, M.
submitted to the Protein Sequence Database, April 1996
A/Reference number: S62975
A/Accession: S69972
A/Molecule type: DNA
A/Residues: 1-1749 <BR>
A/Cross-references: EMBL:Z71330; NID:gl301918; PIDN:CAA95924.1; PID:gl301920
A/Note: biosynthesis of this protein involves a +1 frameshift in the codon for:
C/Genetics:
A/Map position: 14L
A/Mobile element: retrotransposon Ty1.N
C/Superfamily: TyB protein

```

Qy 275 EPSNSHMVKSYSLLFRVTRPGRRFNGMNGINETNENIDVNEELPARRKXKNREDGSEKTF-V 333
Db 1146 PPIINSRQTNS-----SLGGI--GDSNAYTTINS-----KKRSLEDNTEIKV 1185
Qy 334 AQMTVFQKNRELQLLDGEYEVAMQEMECPPSKKX---ATWETILDGKRLPPFFETFSOG 389
Db 1186 SRDTWNTKNMR-----SLEPPSKKRIHLIAAVKAV---KSKIPIT--- 1224
Qy 390 PTLQF--TLRWGTETNDKSTAPIAKPLATRNSESLHQENKFGSVKPTQTIAVKESLTTDDL 447
Db 1225 -TLAYDEAITYNKOIKEKE-----KYIEAVHKE-----VNQLLKMKTWDTDKV 1266
Qy 448 QTRKEKTPNENROKLIIFYOFLYNNVT-----ROOTEARDDLHCP-----WCTLNCRKLY 498
Db 1267 YDRKEID-----PKRVINSFIINRKRGDGTHKARFVARGDIQHPDITDVGSMQSNVHHY 1320
Qy 499 SLLKHLKLCHSRFIYNYVYHFKGARIDVINECYDGSVAGNPODTHRQPGFAFSNGPVK 558
Db 1321 ALMTSLSLALDN---NYVI-----TQLDIS--SAY--LYADIKEELYIRPPHGLMNDKLI 1369
Qy 559 RTPITHILVCRPKETKASMSFLES---EDGEVEQOQRTYSSGHNRKLYFHSDTCLPLRPOE 615
Db 1370 RLKKS---LYGLKQSGANWYETIKSYLIEQCDMEBEVRGWSK---VFKNSQVTICLFVDD 1422
Qy 616 MEVDSEDEKQPEWLREKTIIOIEFSD---VNEGEKEVYKMLWNLHVHMGHGFADNQMHHA 672
Db 1423 MILFSKDLN-----ANKKIITTLKKQYDPTKIINLGE-----SDNEIQYD 1461
Qy 673 CM-LFVENVYGOIKIKNLCRNFMLHLVSMHDFNLISIMSIDKAVT-KLREMOQKLE-KGE 729
Db 1462 ILGLEIKYQSKYMK-----LGMESKSLTEKLPKLVNHLNPKGK 1499
Qy 730 SASPANE-----EITEQNGTANGFPSEINSKEKALETDSVGSVGS 768
Db 1500 KLSAPGQPGLYIDQDELEIDDE-----YKEKVHEMQKLIIGLA 1537

RESULT 8
T38077
hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38077
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z21767
A:Accession: T38077
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1957 <CON>
A:Cross-references: EMBL:Z70690; PIDN:CA94624.1; GSFDB:GN00066; SPDB:SPAC1F3.066
A:Experimental source: strain 972h-; cosmid c1F3
C:Genetics:
A:Gene: SPDB:SPAC1F3.066
A:Map position: 1

A;Molecule type: DNA
A;Residues: 1-1957 <CON>
A;Cross-references: EMBL:Z70690; PIDN:CAA94624.1; GSPDB:GN00066; SPDB:SPAC1F3.06C
A;Experimental source: Strain 972h; cosmid c1F3
C;Genetics:
A;Gene: SPDB:SPAC1F3.06C
A;Map position: 1

Db 991 DIEHLKKNVSEVEVERNALLASNERLMDLKNNGENIASLQTEI---EKKRAENDDLQS 1046
 QY 239 PIRQVPTGKQVPLPD-----LNQTK--FCNFPFLAVSNPEPNSHMAVKSYSL 287
 Db 1047 KLSVVSSEVENLLLSQTNKSLEKTNQKLYEKNQVOKLDEKQDNVELEETSYKYG 1106
 QY 288 LFRVTRPGRREFNGINGETNENIDVNEELPARKKRNREDGE--KTFVAQMTVPDKNRL 345
 Db 1107 L-----GS--ENAIQKDELLAKKSKQCHDLCAAFVDDLK--EKS DAL 1146
 QY 346 QLLDGEYEVAMQEMECIPISKRRATWETILDGKRLPPETFSQGTQFTLRWTGETNDK 405
 Db 1147 EQLTNEKRELIVSLBQ-----SNNNE 1168
 QY 406 STAPIAKPLATNSB---SLHQENKPGSVKPTQTIIVAKESLTTDLQTRKEKDTNENRQK 462
 Db 1169 ALVEERSDLANLSDMKKSLSDSDNVISIRSDLVVNDELDT---LKKDKDS----- 1218
 QY 463 LRIFYQFLYNNNTRQOTEARDDLHCPWCTLNCRKLYSLKHLKLCRGRFIPNYVHPK-- 520
 Db 1219 -----LSTQYSEVCQDRDD-----LLDSLKGCEESF-NKYAVSLREL 1254
 QY 521 --GARIDYSINECYDGSX--AGNPQDIHR-----QPGF----- 549
 Db 1255 CYKSEIDVPVSEILDNFFVAGNFSELSRLTVLSLENYLDAFNQVAFKQKMLNRLTTT 1314
 QY 550 --AFSR-----NGPVKRTPIITHILVCRPK---RTKASMEFLLES-EDGEVEQ 590
 Db 1315 DAFTKVVADLEKLOHEHDDMLIQRDLEKALDKSEKNFLRKEAEMTNIHSLBEGKEET 1374
 QY 591 QRTYSGHNRLYFHSDDTCLPAPQEMEVDSDEKDPWLRK--TIQIIEFSDVNEGEK 648
 Db 1375 KKETAELSRLEDNQLATNKLQNDHLNQBIKEDVLKESLII-SLEP-SLSNQQRK 1433
 QY 649 EVMKLNVLVVKHGFADNMNMHACMLFVENVGQKIKKNI-CRNFMLHLVSMHDFNLIS 707
 Db 1434 ESSLDAKNELEHMLDDTSRKNSSLMEXIESINSLLDDKSFELASAVEKLGALQKLSHS 1493
 QY 708 IMSIDKAVTKLREMOQKLEKESGSA-SPANEBITEQNGTANGPSEINSKEKAL 759
 Db 1494 LSLMENIKSQLEAKEIKQVDESTIQELDHEBITASKN---NYEGLKNDKDSII 1543
 RESULT 9
 B44450
 ubiquitin-specific proteinase (EC 3.4.-.-) UBP3 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YER151c; ubiquitin-specific processing proteinase
 C:Species: Saccharomyces cerevisiae
 C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 21-Jul-2000
 C:Accession: B44450; S50654; S27467
 R:Baker, R.T.; Tobias, J.W.; Varshavsky, A.
 J. Biol. Chem. 267, 23364-23375, 1992
 A:Title: Ubiquitin-specific proteases of Saccharomyces cerevisiae. Cloning of UBP2 and UBP3
 A:Reference number: A44450; MUID:93054674; PMID:1429680
 A:Accession: B44450
 A:Molecule type: DNA
 A:Residues: 1-912 <BAK>
 A:Cross-references: EMBL:M94917; NID:gl73129; PIDN:AAA35191.1; PID:gl73130
 A:Note: sequence extracted from NCBI Backbone (NCBIP:118310)
 R:Dietrich, F.S.
 submitted to the EMBL Data Library, December 1994
 A:Description: The sequence of S. cerevisiae cgmids 8229, 9115, 9132, 9981, and lambda
 A:Reference number: S50654
 A:Accession: S50654
 A:Molecule type: DNA
 A:Residues: 1-912 <DIB>
 A:Cross-references: EMBL:U18917; NID:G603377; PID:G603391; MIPS:YER151c
 C:Genetics:
 A:Gene: SGD:UBP3
 A:Cross-references: SGD:S0000953; MIPS:YER151c
 A:Map position: 5R
 C:Keywords: hydrolase

Query Match 3.7%; Score 150; DB 2; Length 912;
 Best Local Similarity 18.5%; Pred. No. 0.13;
 Matches 153; Conservative 126; Mismatches 287; Indels 260; Gaps 37;
 QY 65 MTDAAARREGESKKIQLPKSLTSLSSVSRGVNPPPHSHSGSLTPVTPPTTPSSSSFRS 124
 Db 3 MODANKEESYSYMK-----TSSPPPTPTNMQIPIYQAPLOMYGTQ 45
 QY 125 STPTETQI--VRLRTRNLI--APIFLHRT-----LTYMSHRN-----SRTNLRKRTF 169
 Db 46 APYLPTQIIPAYSP-----NMVNQNPYI-HQSGSPHILPPQNNINGSGSTNNNNKKW 100
 QY 170 KVDDMLSKYERKMGQESHSLSAHLQLTFTGFFHKNDK-----PSPNS-----E 213
 Db 101 HSGNITNN-NGSSGNGQANSNGSGMSYKSHYTHYNNHIPPMMASPNSSGSGNAGMKQT 159
 QY 214 NEQN-----SVTLVLLVKVCHKRRKDVSCPIRQVPTGKKQVPLIDLNQT 259
 Db 160 NGSNGSSATSPSYSSVNSQYDLYKPDVTKLKN-----LKENSNNLIQPLF--INTT 213
 QY 260 KPG-----NPPSLAVSNBEPFSN-----SHMVKSYSLFRVTRPGRREFNGM 302
 Db 214 EAFPAASVQRYELNMKALNLSSELSSESVKSSAHHHTKSHSI-----PKHNEE 264
 QY 303 INGETN-ENIDVNEELPARKKRNREDGEKTFVAQMTVPDKRRLQLLDGEYEVAMQEMEE 361
 Db 265 VKTETHGEEDAHDKPHASKDAHELKKKTEYK-----EDAKQDRNE 307
 QY 362 CPISKRRATWETILDGKRLPPETFSQGTQFTLRWTGETNDKSTAPIA---KPLATR 417
 Db 308 KVIQEQATVLPVVKKE--PEESVEEN-----TSKTSPPSPSPAASWSAISAD 356
 QY 418 NSESLHQENKPGSVKPTQTIIVAKESLTTDLQTRKEKDTNENRQKLRIFYOFLYNNNTRQ 477
 Db 357 AIKRSQASNKTVSGSVMTKTPISGTTAGVSSTNMAAATIGKSSPL-----LSK 405
 QY 478 QTEARDLHCPWCTLNCRKLYSLKHLKLCRGRFIPNYVHPKGRDIVSINECYDGSYA 537
 Db 406 QKQKDKKVPSTGKIEPLGSA--LRMCPDPDFISVLRNKNQDVENKIPVHSII----- 458
 QY 538 GNPQDIHQPGFAFNRNGPVKRTPIITHILVCRP-----KTKASM----- 577
 Db 459 --PRGIINRANICF-----MSSVLQVLLYCKPFTDVLNSTRNTRNGVGTSSCKLLDA 510
 QY 578 -----SEFLESDG--EVEQORYSSGHNRLYFHSDDTCLPLPQEMEVDS 621
 Db 511 CLUTMYKQFDPKTEYKFFLENADDAKTTESDAKSSKSKSFQHCATADAVKPDF----- 565
 QY 622 DEKDPWLRKEKTIQIEBPSDVNEGEKVMKLNHLVNMKH-----GFIADNMNHA 672
 Db 566 -----YKLTSTIPKFKDLQWGHQDEBEFLTHLLDQLHEELISALDGLTONE--- 612
 QY 673 CMLFVENYQKIIKNNLCRNFMLHVSMMH---DFNLISIMSIDKAVTKLREMOQKLEGE 729
 Db 613 ----IQNMLQSIHQDL-KVFFIRNLSRYKAEF-----IKNASPRKLIELIYK--- 656
 QY 730 SASPANEETIEQNGTANGPSEINSKEKALTDSDSVSGVSKOSKQK 775
 Db 657 --GVINDDSTEB-----NGWHE-----VSGSSKRGKTK 683

RESULT 10
 T18235
 transcription activator GAL11 homolog - yeast (Candida albicans)
 C:Species: Candida albicans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18235
 R:Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, November 1998
 A:Reference number: Z18831
 A:Accession: T18235
 A>Status: preliminary; translated from GB/EMBL/DBJ

A;Cross-references: EMBL:U35149; NTD:g1049248; PID:g1049249; PIDN:AAA81320.1

Query Match 3.6%; Score 147.5; DB 2; Length 1368;
Best Local Similarity 18.7%; Pred. No.0.33;
Matches 174; Conservative 137; Mismatches 324; Indels 297; Gaps 46;

| | | |
|----|-----|--|
| Qy | 33 | IEDNH-----IDTPRVLEKQ---ELCQPTVALSYINRFMTDAARRQE---74 |
| Db | 20 | IDENHPLIKKHVHNLTGSEHRVLLNYGCKUPERYDSKFNVNPKYKNDSSSEESDF79 |
| Qy | 75 | -----SLKKKIQPKLSLTLSSSVSRGNVSTPPRHSSGSLTPVPPTPITPSSSFRST126 |
| Db | 80 | SNLFOGSEFIPKPEPVKNVPLGSK-----NNNSGS-----SFFPSPSPFNDDK122 |
| Qy | 127 | PTEPTQIVAFTRNLAPIFLHRTLVMSHRNSRNRIKTKTFKVDMLSKYKMKGSGQE186 |
| Db | 123 | PTNS-----EILKDDNIEXI---VDATVEDTSGNFNIKREDNDAEKVELKLSGND174 |
| Qy | 187 | S-HSLV-----AHLQITFTGFHKNDKPSNSENQNSVTLLEVLLV--K227 |
| Db | 175 | SINNLISINNIIDNTEHKVQVQNEQEIESKSFNDND---NSINDISS---VLLI126 |
| Qy | 228 | VCHKKRKOVSCEIRQVPVKKQVPLIPDLNQTPGPFPSLAYSSNE-PEPS-----NSHM281 |
| Db | 227 | TEHKVQVQNEQEIESKSFNDNDNSINDITENKEH--KVQVQNEQEIESKLSGND284 |
| Qy | 282 | VKSYSLLFRVTRPGRREFNGMINGETNENIDV--NEELPARRKRNREDGEKTFVQM2VF339 |
| Db | 285 | INDISI-----NNVIDNTEHKVQVQNEQEIESKSFNDNDNSINDISIMILLI332 |
| Qy | 340 | D-----KURELQ-----LLDGEYEVAMQWMECPISKKRATWEIL375 |
| Db | 333 | DNETEHKVQVQNEQEIESKLSGNDLNNVNIENKEQIVQVEKKASRRNM-----387 |
| Qy | 376 | DGKRLPPPTFSQGTLOF-----TLRWGTETN-403 |
| Db | 388 | -----IPDEIRQGVIKKYKSSNNHIIKMNIIVEKHEYLTSQSEYKQKLDMLRNNRVNL441 |
| Qy | 404 | ----DKSTAPIAKPLATNSLSLH---QEN-----KESGKVPQTIAVKESELT--TDLQT449 |
| Db | 442 | FLINKKNDL-KPNEKH-EYIHVSEENEQIVKEIQLKPLQLIKVNENIMFNNTN499 |
| Qy | 450 | RKEKTPNENRQKLIIFYQFLYNNNTNQTEARDDLHCPWCTLNCRLKYS-----LLKHLK505 |
| Db | 500 | NDTKBEYTDNMNEVSPNTOLYESNMENYQNRG-----IYNCDQIPDSEVCIPKQE550 |
| Qy | 506 | LCHSRFIF-----NVVYHPKGARIDVSNICYDGSVAGNPQDIHROPGPFAFSRNGPV557 |
| Db | 551 | YFHNNEIYQLNGIFPGSEVISPQK---EYSYNDIYQLYQMTPNNEIYIPTKGYFENNEI607 |
| Qy | 558 | KRTP--ITHILVCRPK-----RTKASNSEFLESDGEVEBQQ591 |
| Db | 608 | YLSDDMLSDCEVCTPKYBIEQNEBIYPLNGIFGSDNEIYITEQKGLDKIEYPLTVGFPSDN667 |
| Qy | 592 | RTYSSG-----NRLYFHSDDTCLPLR---PQMEYDSDSEKDP626 |
| Db | 668 | EYVITEHTGSLNEECPLNEVLLDNEIITQIGSHDEEYPLNEILPNEVITEQDGS---725 |
| Qy | 627 | EWLREKTTITBEFSDVNEG-----BEKVKNLWNLHVMMKHGFIADNQNHACMLFV677 |
| Db | 726 | --LDDEIYPLTGVFSD-NETVITEQKGLDKIEYPL-----TGVSFDSNE---TYIT770 |
| Qy | 678 | ENVYQKIIKKNLCRNFMHLHYSMDHFNLS--IWSIDKAVTKLREMQCKLEKG---ESAS732 |
| Db | 771 | EHGS--LNEEIC---PLNEVLL--DNBIITEQIGSLDEAICPLNEILLDNEILSTLKEKDS824 |
| Qy | 733 | PANBEITBEQNGTANGFSEINSKKALETDSV764 |
| Db | 825 | QNEI1CTPKQEVSON--EEINTLNEILSDNEI854 |

TyB protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty1.P

N/Alternate names: protein P9659_6
C/Species: *Saccharomyces cerevisiae*
C/Date: 09-Mar-1996 #sequence_revision 06-Sep-1996 #text_change 23-Mar-2001
C/Accession: S69980
R/Fulton, L.
submitted to the EMBL Data Library, November 1995
A/Description: the sequence of *S. cerevisiae* cosmid 9659.
A/Reference number: S69022
A/Accession: S69980
A/Molecule type: DNA
A/Residues: 1-1755 <FUL>
A/Cross-references: EMBL:U40829
A/Note: biosynthesis of this protein involves a +1 frameshift in the codon for residue 41:
C/Genetics:
A/Map position: 16R
A/Mobile element: retrotransposon Ty1.P
C/Superfamily: TyB protein

Query Match 3.6%; Score 147.5; DB 2; Length 1755;
Best Local Similarity 22.5%; Pred. No. 0.46;
Matches 158; Conservative 98; Mismatches 234; Indels 213; Gaps 44;

Qy 38 ITDPRVL---EKQLQOPTYVALSY---INRFMTD-----AARQESLKKKI----- 80
Db 881 VDTNTYVILQKESRLDQFNFDALTFDELRLTASYSFASIEQIESDNLNIESDHDF 940
Qy 81 -----OPKLSLTGLASSYGRGNVSTP-----RHSSGSLTP--VTPITPS-- 119
Db 941 QSDIELHPQPR--NVLSKAVSPTD--STPPSTHTEDSKSVKTNIRAPREVDPNISESNI 997
Qy 120 ---SSPRSPTEPTQYRFLTRN-----LTAPIFLHRTITYMHSRNSRNIRKKT 168
Db 998 LPSKKRSSTP---QISNIESTGSGGMHKNVPLLAPMSQNT-----HESSHAS--KSKD 1047
Qy 169 FKVDDMLSKVMKMEQESHLSLAHLQTTFGFFHKNDKPSNPSENGSVTLVLLVKV 228
Db 1048 FRHSOSYS-----ENETNHTNVPISSTG--GTNNKTPQISDQTE-----KRI 1089
Qy 229 CHKGRK-DVS-----CFIROVPTGKKQ-----VPLIPDLNQTQKGNFPFLA 268
Db 1090 IHRSPSIDASPENNSSHNIVPKPTTVSQNGTEESIIADLPL-PDLPSPSTPEFPD-- 1146
Qy 269 VSNFEFPNSHMVKSYSLLFVTFPCRRPENGINGETNENIDVNEELPARRKNEEDG 328
Db 1147 -PFKELPPTINSHTNS-----SLGGI--GDSNAYTTINS-----KKRSLEDN 1185
Qy 329 EKTFF-VAQMTVPDKRRLLQLLDGEYEVAMQMEECPSKKR---ATWETILDGRLPPF 383
Db 1186 ETEIKVSRDTRWNTKNWR-----SLEPPRSKKRIHLIAAVKAV--KSIKPI 1228
Qy 384 EFTSQGPTLQF--TLRWGTGETNDKSTAPIAKPLATRNSESLSHQENKGSVKPTQTIAVKE 441
Db 1229 RT-----TLRYDEATYNAKTIKEK-----KYIEAYHKE-----VNOLLKNT 1266
Qy 442 SLTTDLQTRKEKDTENENRQKLRIFYOFLYNNNT-----RQOTEARDDDLHCF-----WCTL 492
Db 1267 WOTDKYDRKEID-----PKRVINSMFINKKRDGTHKARFVARDGIQHDPDYDTGMQS 1320
Qy 493 NCRKLYSLKLKHLKCHSRPIFNVYVHPKGARIDVINECYDGSYAGNPQDTHRQGFAGFS 552
Db 1321 NTWHYALMTSLSLADN---NYYI---TQLD-S--SAY--LYADIKEELYIEPPPHLG 1369
Qy 553 RNGPVKRTPIITHLCRPKRTKASMEFLES-----EDGEVEQORTYSSGHNRLPHSPTCL 609
Db 1370 MDDKLIRLKSXYGL---KQSGANWYEIYKSLYIKQCGMEVVRGWSG-----VPKNSQVTI 1422
Qy 610 PLRPQMEVDSDEKDPBWLREKTTITQTEBPSD---VNEGEKE 649
Db 1423 CLFVDDMILFSKDLN---ANKKIITTLKKQYDTKINLGESD 1461

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:57:34 ; Search time 138 Seconds
(without alignments)
1769.128 Million cell updates/sec

Title: US-09-874-162A-8
Perfect score: 4078
Sequence: 1 MTGIAAASFFSNTCRFGCGG.....KALETDSVSGVSKSKQKL 776

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1297172 seqs, 314612898 residues

Total number of hits satisfying chosen parameters: 1297172

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 4078 | 100.0 | 776 | 9 | US-09-874-162A-8 |
| 2 | 3413 | 83.7 | 739 | 9 | US-09-874-162A-5 |
| 3 | 2043 | 50.1 | 388 | 9 | US-09-764-864-931 |
| 4 | 1497 | 36.7 | 289 | 9 | US-09-764-864-911 |
| 5 | 1453 | 35.6 | 292 | 9 | US-09-764-864-1369 |
| 6 | 919 | 22.5 | 175 | 9 | US-09-764-864-1355 |
| 7 | 661 | 16.2 | 243 | 9 | US-09-874-162A-2 |
| 8 | 661 | 16.2 | 243 | 15 | US-10-144-198-16 |
| 9 | 661 | 16.2 | 243 | 15 | US-10-104-047-2404 |
| 10 | 652 | 16.0 | 278 | 9 | US-09-764-864-1334 |
| 11 | 476 | 11.7 | 143 | 9 | US-09-764-864-884 |
| 12 | 359 | 8.8 | 69 | 14 | US-10-029-386-28495 |
| 13 | 280 | 6.9 | 171 | 12 | US-10-276-774-2060 |
| 14 | 214.5 | 5.3 | 813 | 14 | US-10-231-778-2 |
| 15 | 211 | 5.2 | 692 | 14 | US-10-177-478-6 |

ALIGNMENTS

RESULT 1

US-09-874-162A-8
; Sequence 8, Application US/09874162A
; Patent No. US20020155452A1
; GENERAL INFORMATION:
; APPLICANT: Koontz, Jason
; APPLICANT: Sklar, Jeffrey
; TITLE OF INVENTION: FUSION OF JAZF1 AND JJAZ1 GENES IN
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS
; FILE REFERENCE: 05311-024001
; CURRENT APPLICATION NUMBER: US/09/874,162A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,093
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-162A-8

| | | | | | | | |
|-----------------------|---------|--|-------|------------|----|--------|------|
| Query Match | 100.0%; | Score | 4078; | DB | 9; | Length | 776; |
| Best Local Similarity | 100.0%; | Pred. No. | 0; | | | | |
| Matches | 776; | Conservative | 0; | Mismatches | 0; | Indels | 0; |
| Gaps | 0; | | | | | | |
| Qy | 1 | MTGIAAASFFSNTCRFGCGGLHPTLADLIEHIEDNHIDTPRVLEKQELQQPTVVALSY | 60 | | | | |
| Db | 1 | MTGIAAASFFSNTCRFGCGGLHPTLADLIEHIEDNHIDTPRVLEKQELQQPTVVALSY | 60 | | | | |
| Qy | 61 | INRMTDAARREQESLKKIKQPKLSLTSSVSGRGNVSTPPRHSSGSLTPTVPTPSS | 120 | | | | |
| Db | 61 | INRMTDAARREQESLKKIKQPKLSLTSSVSGRGNVSTPPRHSSGSLTPTVPTPSS | 120 | | | | |
| Qy | 121 | SFRSSTTEPTQIYRFLRNLNLAIPFLHRTLTVMHSRNTNKRKTFKVDMLSKVEK | 180 | | | | |
| Db | 121 | SFRSSTTEPTQIYRFLRNLNLAIPFLHRTLTVMHSRNTNKRKTFKVDMLSKVEK | 180 | | | | |

QY 181 MKGEQSHLSAHLQITFTGPHKNDKSPNSNEQNSVTLEVLVVKVCHKRRKDVSCPI 240
 Db 181 MKGEQSHLSAHLQITFTGPHKNDKSPNSNEQNSVTLEVLVVKVCHKRRKDVSCPI 240
 QY 241 ROVPTGKQVPLIPDLNQTGKGNFPLAVSSNEFEPNSHVMVKSYSLLFRVTRGRRFEN 300
 Db 241 ROVPTGKQVPLIPDLNQTGKGNFPLAVSSNEFEPNSHVMVKSYSLLFRVTRGRRFEN 300
 QY 301 GMINGETHENIDVNEELPARRKRNREDGKTFVAQMTVFDKRRRLQLLDGGEYVAMQEME 360
 Db 301 GMINGETHENIDVNEELPARRKRNREDGKTFVAQMTVFDKRRRLQLLDGGEYVAMQEME 360
 QY 361 ECPISKKAATWETILDGKRLPPFETFSQGPLTQTLRWGTGETNDKSTAPIAKPLATRNSE 420
 Db 361 ECPISKKAATWETILDGKRLPPFETFSQGPLTQTLRWGTGETNDKSTAPIAKPLATRNSE 420
 QY 421 SLHQENKFGSVKPTQTIAVKESLATDLOTRKEKOTPNENRQKLRIFYQFLYNNNTRQOTE 480
 Db 421 SLHQENKFGSVKPTQTIAVKESLATDLOTRKEKOTPNENRQKLRIFYQFLYNNNTRQOTE 480
 QY 481 ARDDLHCPCWCTLNCBKLKSLKHLKCHSRFIENYVYHPKGRIDVSNIECYDGSYAGNP 540
 Db 481 ARDDLHCPCWCTLNCBKLKSLKHLKCHSRFIENYVYHPKGRIDVSNIECYDGSYAGNP 540
 QY 541 QDIHRQPGFAFSRNGPVKRTPIITHILVCRPKRTKASMEFLESEDEGEVEQORTYSSGHN 600
 Db 541 QDIHRQPGFAFSRNGPVKRTPIITHILVCRPKRTKASMEFLESEDEGEVEQORTYSSGHN 600
 QY 601 LYFHSDDTCLPLRQPMEDVSEDEKDPWLRKTTQIIEFSDVNEGEKEVMKLNHLVHMVK 660
 Db 601 LYFHSDDTCLPLRQPMEDVSEDEKDPWLRKTTQIIEFSDVNEGEKEVMKLNHLVHMVK 660
 QY 661 HGFIADNOMHACMLFVENYQKIIKKNLCRNFMLHLSVHDFNLISIMSDIKAVTKLRE 720
 Db 661 HGFIADNOMHACMLFVENYQKIIKKNLCRNFMLHLSVHDFNLISIMSDIKAVTKLRE 720
 QY 721 MQKLEKESASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776
 Db 721 MQKLEKESASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776

RESULT 2
 US-09-874-162A-5
 ; Sequence 5, Application US/09874162A
 ; Patent No. US20020155452A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Koontz, Jason
 ; APPLICANT: Sklar, Jeffrey
 ; TITLE OF INVENTION: FUSION OF JAZF1 AND JAZF1 GENES IN
 ; FILE REFERENCE: 05311-024001
 ; CURRENT APPLICATION NUMBER: US/09/874.162A
 ; PRIOR FILING DATE: 2001-06-04
 ; PRIOR FILING DATE: 2000-06-02
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 739
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-874-162A-5

Query Match 83.7%; Score 3413; DB 9; Length 739;
 Best Local Similarity 99.8%; Pred. No. 6.1e-282;
 Matches 647; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 129 EPTQIYRFLRNLIAPIFLHRTLYMSHRNSRNIKRKTFKVDMDLSKVKMKMGQESH 188
 Db 92 KPTQIYRFLRNLIAPIFLHRTLYMSHRNSRNIKRKTFKVDMDLSKVKMKMGQESH 151
 QY 189 SLSAHLQITFTGPHKNDKSPNSNEQNSVTLEVLVVKVCHKRRKDVSCPIQVPTGKK 248

Db 152 SLSAHLQITFTGPHKNDKSPNSNEQNSVTLEVLVVKVCHKRRKDVSCPIQVPTGKK 211
 QY 249 QVPLIPDLNQTGKGNFPLAVSSNEFEPNSHVMVKSYSLLFRVTRGRRFENGMINGETN 308
 Db 212 QVPLIPDLNQTGKGNFPLAVSSNEFEPNSHVMVKSYSLLFRVTRGRRFENGMINGETN 271
 QY 309 ENIDVNEELPARRKRNREDGKTFVAQMTVFDKRRRLQLLDGGEYVAMQEMECCPISKRR 368
 Db 272 ENIDVNEELPARRKRNREDGKTFVAQMTVFDKRRRLQLLDGGEYVAMQEMECCPISKRR 331
 QY 369 ATWETILDGKRLPPFETFSQGPLTQTLRWGTGETNDKSTAPIAKPLATRNSESLHQENK 428
 Db 332 ATWETILDGKRLPPFETFSQGPLTQTLRWGTGETNDKSTAPIAKPLATRNSESLHQENK 391
 QY 429 GSVKPTQTIAVKESLATDLOTRKEKOTPNENRQKLRIFYQFLYNNNTRQOTEARDDLHCP 488
 Db 392 GSVKPTQTIAVKESLATDLOTRKEKOTPNENRQKLRIFYQFLYNNNTRQOTEARDDLHCP 451
 QY 489 WCTLNCRKLYSLKHLKCHSRFIENYVYHPKGRIDVSNIECYDGSYAGNPDIHRQPG 548
 Db 452 WCTLNCRKLYSLKHLKCHSRFIENYVYHPKGRIDVSNIECYDGSYAGNPDIHRQPG 511
 QY 549 FAFSRNGPVKRTPIITHILVCRPKRTKASMEFLESEDEGEVEQORTYSSGHNRLYFHS 608
 Db 512 FAFSRNGPVKRTPIITHILVCRPKRTKASMEFLESEDEGEVEQORTYSSGHNRLYFHS 571
 QY 609 LPLRQPMEDVSEDEKDPWLRKTTQIIEFSDVNEGEKEVMKLNHLVHMKGFIADNQ 668
 Db 572 LPLRQPMEDVSEDEKDPWLRKTTQIIEFSDVNEGEKEVMKLNHLVHMKGFIADNQ 631
 QY 669 MNHACMLFVENYQKIIKKNLCRNFMLHLSVHDFNLISIMSDIKAVTKLREMQCKLEKG 728
 Db 632 MNHACMLFVENYQKIIKKNLCRNFMLHLSVHDFNLISIMSDIKAVTKLREMQCKLEKG 691
 QY 729 ESASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776
 Db 692 ESASPANEETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 739

RESULT 3
 US-09-764-864-931
 ; Sequence 931, Application US/09764864
 ; Patent No. US20020132753A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT223
 ; CURRENT APPLICATION NUMBER: US/09/764.864
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 1792
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 931
 ; LENGTH: 388
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-764-864-931

Query Match 50.1%; Score 2043; DB 9; Length 388;
 Best Local Similarity 99.5%; Pred. No. 2e-165;
 Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 389 GPTLOFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLATDLO 448
 Db 1 GPTLOFTLRWTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLATDLO 60
 QY 449 TRKEKOTPNENRQKLRIFYQFLYNNNTRQOTEARDDLHCPWCTLNCBKLKSLKHLKCH 508
 Db 61 TRKEKOTPNENRQKLRIFYQFLYNNNTRQOTEARDDLHCPWCTLNCBKLKSLKHLKCH 120
 QY 509 SRFIENYVYHPKGRIDVSNIECYDGSYAGNPDIHRQPGFAFSRNGPVKRTPIITHILVC 568
 Db 121 SRFIENYVYHPKGRIDVSNIECYDGSYAGNPDIHRQPGFAFSRNGPVKRTPIITHILVC 180


```
QY 569 RPKRTKASMEFLESDGEVEQRTYSSGHNRLYFHSHTCLPLRPOEVEVDSDEKDEW 628
Db 181 RPKRTKASMEFLESDGEVEQRTYSSGHNRLYFHRDTCLPLRPOEVEVDSDEKDEW 240
QY 629 LREKTIQTIEEFSVDNEGEKVMKLNHLVHMKHGFADNQMNHACMLFVENYQKIIKEN 688
Db 241 LREKTIQTIEEFSVDNEGEKVMKLNHLVHMKHGFADNQMNHACMLFVENYQKIIKEN 300
QY 689 LCRNFMHLVSHMDFNLISIMSIDKAVTKLRMOQKLEKGSASPANBEITEEQNGTANG 748
Db 301 LCRNFMHLVSHMDFNLISIMSIDKAVTKLRMOQKLEKGSASPANBEITEEQNGTANG 360
QY 749 FSEINSKEKALETDSVSGVSKQSKQKL 776
Db 361 FSEINSKEKALETDSVSGVSKQSKQKL 388

RESULT 4
US-09-764-864-911
; Sequence 911, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 911
; TYPE: PRT
; LENGTH: 289
; ORGANISM: Homo sapiens
; US-09-764-864-911

Query Match 36.7%; Score 1497; DB 9; Length 289;
Best Local Similarity 99.6%; Pred. No. 5.2e-119;
Matches 283; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 493 NCRKLYSLKHLKLGHSRIENYVHPKGRADIVSINECYDGSYAGNQDTHRQGFAPS 552
Db 6 DCRKLYSLKHLKLGHSRIENYVHPKGRADIVSINECYDGSYAGNQDTHRQGFAPS 65
QY 553 RNPVKRTPITHLVCRPRTKASMEFLESDGEVEQRTYSSGHNRLYFHSHTCLPLR 612
Db 66 RNPVKRTPITHLVCRPRTKASMEFLESDGEVEQRTYSSGHNRLYFHSHTCLPLR 125
QY 613 PQMEVDSDEKDPWLREKTTIQIEEFSVDNEGEKVMKLNHLVHMKHGFADNQMHA 672
Db 126 PQMEVDSDEKDPWLREKTTIQIEEFSVDNEGEKVMKLNHLVHMKHGFADNQMHA 185
QY 673 CMLFVENYQKIIKNLCNFMHLVSHMDFNLISIMSIDKAVTKLRMOQKLEKGSAS 732
Db 186 CMLFVENYQKIIKNLCNFMHLVSHMDFNLISIMSIDKAVTKLRMOQKLEKGSAS 245
QY 733 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 776
Db 246 PANEEITEEQNGTANGFSEINSKEKALETDSVSGVSKQSKQKL 289

RESULT 5
US-09-764-864-1369
; Sequence 1369, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
```

```
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1369
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (9)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (31)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (46)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (121)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (283)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1369

Query Match 35.6%; Score 1453; DB 9; Length 292;
Best Local Similarity 96.8%; Pred. No. 3e-115;
Matches 275; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 233 RKDVSCPIROVPTGKQVPLIPDLNOKPGNFFPSLAVSSNEFSPNSHVMKYSLLPRVT 292
Db 1 KDVSCPIXQVPTGKQXPLNPLNPLNPLNPLNPLNPLNPLNPLNPLNPLNPLNPLNPLN 60
QY 293 RPSRRFNGMNGTNETNENIDVNEELPARRKRNREDGKTFVAQMTVPDKNRRLQLLDGEY 352
Db 61 RPSRRFNGMNGTNETNENIDVNEELPARRKRNREDGKTFVAQMTVPDKNRRLQLLDGEY 120
QY 353 EVAMQEMEECPISKRRATWETILDGKLPFPFETFSOGPTLQFTLRWTGETNDKSTAPIAK 412
Db 121 XVAMQEMEECPISKRRATWETILDGKLPFPFETFSOGPTLQFTLRWTGETNDKSTAPIAK 180
QY 413 PLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN 472
Db 181 PLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEKDTPNENRQKLRIFYQFLYN 240
QY 473 NNTRQOTEARDDHLHCPWCTLNCRKLYSLKHLKLGHSRIENYV 516
Db 241 NNTRQOTEARDDHLHCPWCTLNCRKLYSLKHLKLGHSRIENYV 284

RESULT 6
US-09-764-864-1355
; Sequence 1355, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1355
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (166)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1355

Query Match 22.5%; Score 919; DB 9; Length 175;
Best Local Similarity 94.3%; Pred. No. 5.2e-70;
Matches 165; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 466 FYQFLYNNNTROQTEARDLHCPWCTLNCRLKLSLLKHLKLSRFFINYYHYPKGAID 525
DB 1 FYQXLYNNNTROQTEARDLHCPWCTLNCRLKLSLLKHLKLSRFFINYYHYPKGAID 60
QY 526 VSINECYDGSVAGNPDHROPGFAPSRNGPVKRTPIITHILVCRPKTKASMSFFLESED 585
DB 61 VSINECYDGSVAGNPDHROPGFAPSRNGPVKRTPIITHILVCRPKTKASMSFFLESED 120
QY 586 GEVEQRTYSSGHNRLYFHSSTCLPLRPQEMVDSDEKDPWLRKXKPLHLKXSF 640
DB 121 GEVEQRTYSSGHNRLYFHSSTCLPLRPQEMVDSDEKDPWLRKXKPLHLKXSF 175

RESULT 7
US-09-874-162A-2
; Sequence 2, Application US/09874162A
; Patent No. US2002015542A1
; GENERAL INFORMATION:
; APPLICANT: Koontz, Jason
; APPLICANT: Sklar, Jeffrey
; TITLE OF INVENTION: FUSION OF JAZF1 AND JAZ1 GENES IN
; TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS
; FILE REFERENCE: 05311-024001
; CURRENT APPLICATION NUMBER: US/09/874,162A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,093
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-162A-2

Query Match 16.2%; Score 661; DB 9; Length 243;
Best Local Similarity 100.0%; Pred. No. 9.1e-48;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTGIAAASFFSNTCRFGGCGLHFFPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
DB 1 MTGIAAASFFSNTCRFGGCGLHFFPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
QY 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
DB 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
QY 121 SFRSSTPT 128
DB 121 SFRSSTPT 128

RESULT 8
US-10-144-198-16
; Sequence 16, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies Inc

; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-16

Query Match 16.2%; Score 661; DB 15; Length 243;
Best Local Similarity 100.0%; Pred. No. 9.1e-48;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTGIAAASFFSNTCRFGGCGLHFFPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
DB 1 MTGIAAASFFSNTCRFGGCGLHFFPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
QY 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
DB 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
QY 121 SFRSSTPT 128
DB 121 SFRSSTPT 128

RESULT 9
US-10-104-047-2404
; Sequence 2404, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2404
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2404

Query Match 16.2%; Score 661; DB 15; Length 243;
Best Local Similarity 100.0%; Pred. No. 9.1e-48;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTGIAAASFFSNTCRFGGCGLHFFPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
DB 1 MTGIAAASFFSNTCRFGGCGLHFFPTLADLIEHIEDNHIDTDPVLEKQELQOPTYVALSY 60
QY 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
DB 61 INRFMTDAARREQESLKKKIQPKLSLTLSVSVSRGNVSTPPRHSSGSLTPPVTPPTPSS 120
QY 121 SFRSSTPT 128
DB 121 SFRSSTPT 128

RESULT 10
US-09-764-864-1334
; Sequence 1334, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1334
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1334

Query Match      16.0%; Score 652; DB 9; Length 278;
Best Local Similarity 99.2%; Pred. No. 6.6e-47;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTPRVLEKQLQOPTVALSY 60
Db 36 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTPRVLEKQLQOPTVALSY 95

QY 61 INFMTDAARRQESLKXKIQPKLSLTLSVVSRGNVSTPPRHSSGSLTPPVTPITPSS 120
Db 96 INFMTDAARRQESLKXKIQPKLSLTLSVVSRGNVSTPPRHSSGSLTPPVTPITPSS 155

QY 121 SPRSSTPT 128
Db 156 SPRSSTPT 163

RESULT 11
US-09-764-864-884
; Sequence 884, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 884
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-884

Query Match      11.7%; Score 476; DB 9; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.5e-32;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTPRVLEKQLQOPTVALSY 60
Db 51 MTGIAAASFNTCRFGGGLHPTLADLIEHEDNHIDTPRVLEKQLQOPTVALSY 110

QY 61 INFMTDAARRQESLKXKIQPKLSLTLSVVSRGNVSTPPRHSSGSLTPPVTPITPSS 93
Db 111 INFMTDAARRQESLKXKIQPKLSLTLSVVSRGNVSTPPRHSSGSLTPPVTPITPSS 143

RESULT 12
US-10-029-386-28495
; Sequence 28495, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
```

```

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28495
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR17.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P23477, EVALUATE 2.70e+00
US-10-029-386-28495

Query Match      8.8%; Score 359; DB 14; Length 69;
Best Local Similarity 98.6%; Pred. No. 7.9e-23;
Matches 68; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 569 RPXETKASMEFLESEDEGEVQQTYSXSSGHNRLYFHSDTCLPLRPQMEVDSDEKDPW 628
Db 1 RPXETKASMEFLESEDEGEVQQTYSXSSGHNRLYFHSDTCLPLRPQMEVDSDEKDPW 60

QY 629 LREXITITQI 637
Db 61 LREXITITVI 69

RESULT 13
US-10-276-774-2060
; Sequence 2060, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2060
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-2060

Query Match      6.9%; Score 280; DB 12; Length 171;
Best Local Similarity 100.0%; Pred. No. 1.8e-15;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 QESLKKKIQPKLSLTLSVVSRGNVSTPPRHSSGSLTPPVTPITPSSFRSSTPT 128
Db 1 QESLKKKIQPKLSLTLSVVSRGNVSTPPRHSSGSLTPPVTPITPSSFRSSTPT 56

RESULT 14
US-10-231-778-2
; Sequence 2, Application US/10231778
; Publication No. US20030126647A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau, Pierre
; APPLICANT: Chaudhury, Abdul M.
```



```

QY 341 KNRRLQLDGEYEVAVQWECPISKKRATWETILDGKRLPPFFTFSGPTLQFTLRWTG 400
Db 392 SS-----EPKVRHVNDNVSS-----TPRAHS 413
QY 401 ETNDKSTAFIAKPLATRNSESLHQENKPGSVKPTQTIKVESLTTDLQTRKEKDTPENR 460
Db 414 SKKNKST-----RKNDNIPSPPKTRSSKKTNLTTR---TOPAIAESEPKVPHVND 462
QY 461 QKLRIFYOFLYNNTRQOTEARDD-LHCFWCTLNCRKLYSLKHLKLCCHSRFIFNYVYHP 519
Db 463 DKVSSTPR-AHSSKKNKSTHKDDNASLPPKTRSSKKTSDIL----- 503
QY 520 KGARIDVSIKNECYDGSYAGNPQDIHRQPGAFSRNGPVKRTPTITHILVCRPKRTKASME 579
Db 504 -----ATTQPAKAP-----SEPKVTRVSRK 525
QY 580 FLESDGEVEQOQTYSSGHNRLYFHSDTCLPLRQEM--EVDSEDEKDPFWLREKTIITQI 637
Db 526 ELHAERCEAKRLERLK--GRQFYHSQTMQPTFEQVMSNEDSENETDDYALDI-SERLRL 582
QY 638 EEFSDVNEGEKEVMKLNHLVNMKGFIADNQMHACMLFVE-----NYGQKLI 685
Db 583 ERLVGVSKBEKYMVLMNIFVKQKRVADGHVPWACEBFKHLKHEEMKNSSSFDWWWMF 642
QY 686 KKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMQQKLEKGESASPANEEITE 740
Db 643 RIKLWNGLICAKTFHKCTTILLSNDEA-----GQFTSGSAANANNOQSME 689

```

Search completed: August 25, 2004, 18:09:35
Job time : 148 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 25, 2004, 17:53:09 ; Search time 34 Seconds
(without alignments)
1178.287 Million cell updates/sec

Title: US-09-874-162A-8
Perfect score: 4078
Sequence: 1 MTGIAAASFFNTCRFGCGG.....KALETDSVSGVSKSKQKL 776

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A-COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B-COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A-COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B-COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS-COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1 | 238.5 | 5.8 | 604 | US-09-339-947A-8 | Sequence 8, Appli |
| 2 | 224 | 5.5 | 611 | US-09-339-947A-1 | Sequence 1, Appli |
| 3 | 159 | 3.9 | 1494 | US-08-755-587-186 | Sequence 186, App |
| 4 | 156.5 | 3.8 | 3696 | US-09-134-001C-5080 | Sequence 5080, Ap |
| 5 | 150 | 3.7 | 912 | US-07-789-915A-8 | Sequence 8, Appli |
| 6 | 150 | 3.7 | 912 | US-08-005-002C-8 | Sequence 8, Appli |
| 7 | 150 | 3.7 | 912 | US-08-487-203A-8 | Sequence 8, Appli |
| 8 | 146.5 | 3.6 | 1235 | US-08-118-101A-2 | Sequence 2, Appli |
| 9 | 142.5 | 3.5 | 1972 | US-09-418-710-21 | Sequence 21, Appli |
| 10 | 137.5 | 3.4 | 2954 | US-09-150-867-1 | Sequence 1, Appli |
| 11 | 135 | 3.3 | 3248 | US-08-353-700-1 | Sequence 1, Appli |
| 12 | 135 | 3.3 | 3248 | PCT-US95-16216-1 | Sequence 1, Appli |
| 13 | 132 | 3.2 | 2482 | US-08-328-254-6 | Sequence 6, Appli |
| 14 | 128.5 | 3.2 | 907 | US-08-938-830-26 | Sequence 26, Appli |
| 15 | 128.5 | 3.2 | 907 | US-09-020-222-26 | Sequence 26, Appli |
| 16 | 128.5 | 3.2 | 1167 | US-09-803-671B-2 | Sequence 2, Appli |
| 17 | 128 | 3.1 | 798 | US-09-203-453-5 | Sequence 5, Appli |
| 18 | 128 | 3.1 | 798 | US-09-236-5 | Sequence 5, Appli |
| 19 | 128 | 3.1 | 1164 | US-09-457-708-2 | Sequence 2, Appli |
| 20 | 128 | 3.1 | 1164 | US-09-950-046A-2 | Sequence 2, Appli |
| 21 | 128 | 3.1 | 1164 | US-09-976-594-989 | Sequence 989, App |
| 22 | 127.5 | 3.1 | 609 | US-09-107-532A-4451 | Sequence 4451, Ap |
| 23 | 126.5 | 3.1 | 3418 | US-08-639-501-2 | Sequence 2, Appli |
| 24 | 126.5 | 3.1 | 3418 | US-08-603-753D-4 | Sequence 4, Appli |
| 25 | 126.5 | 3.1 | 3418 | US-09-044-946-2 | Sequence 2, Appli |
| 26 | 126.5 | 3.1 | 3418 | US-08-755-587-44 | Sequence 44, Appli |
| 27 | 126.5 | 3.1 | 3418 | US-09-044-908-2 | Sequence 2, Appli |

ALIGNMENTS

RESULT 1
US-09-339-947A-8
; Sequence 8, Application US/09339947A
; Patent No. 6630616
; GENERAL INFORMATION:
; APPLICANT: YOSHIDA, No. 6630616umasa
; APPLICANT: YANAI, Yukihiko
; APPLICANT: KATO, Yoshihiro
; APPLICANT: HIRATSUKA, Junzo
; APPLICANT: TAKAHASHI, Shigeru
; APPLICANT: MIWA, Tatsuehi
; TITLE OF INVENTION: GENE FOR FLORAL REGULATION AND METHODS FOR CONTROLLING
; FILE REFERENCE: 032735-001
; CURRENT APPLICATION NUMBER: US/09/339,947A
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: JP 1998-180065
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: JP 1999-179043
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: ZN FING
; LOCATION: (310)..(335)
US-09-339-947A-8

Query Match 5.8%; Score 238.5; DB 4; Length 604;
Best Local Similarity 18.9%; Pred. No. 2.5e-13;
Matches 130; Conservative 95; Mismatches 242; Indels 221; Gaps 22;

| | | | |
|----|-----|---|-----|
| QY | 129 | BPTQYRFLRNRNLIADIFLHRTLYMSH--RNSRTNI-----KKTFFKVDMLSKVE | 179 |
| DB | 27 | KPVLYNIIQERSKNPAFLQRCCLLYKHARKKRSLLTISLSGGTNKELRAQNIFFLYV | 86 |
| QY | 180 | KMKGEQSHSLSAH-----LQLTFTGFFHKNDKPS-----PNSGN--EQNSVTLEV | 223 |
| DB | 87 | LLARPNTNVSLEGHSPYRFRSACLTSFHEFGNKDYTEATFVDPVKNLATSACSLNI | 146 |
| QY | 224 | LLVKVCHKKR--DVSCPIRQVPTGKKQ-----VPLIDNLQTKPGNFPPLAVS | 270 |
| DB | 147 | ILIS-CGRAEOTTFDDNNCSGNHVEGSTLQLEKCFGKGIPI--DLLASSGNCVLSL- | 202 |
| QY | 271 | SNEPEPSNHVKSYSLLFRVTRPGRREFNGMETNEMIDVNEELPA--RRKRNREDG | 328 |
| DB | 203 | -----GHTVEMSSVTVENTPSPLEPKFLEDD | 228 |

APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 186:
SEQUENCE CHARACTERISTICS:
LENGTH: 1494 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-755-587-186

Query Match 3.9%; Score 159; DB 3; Length 1494;
Best Local Similarity 18.3%; Pred. No. 3e-05;
Matches 177; Conservative 151; Mismatches 355; Indels 284; Gaps 46;
QY 23 EPTLADLIEHEDNHDTPVLEKQELQOPTYVA-LSVINRFMTDAARREOESLKKIKQ 81
DB 370 FTASKEIKLSNNIRKSMFLKDIIEHYPTNLACLEIVNTSSLESOKKPSK--HALD 427
QY 82 PKLSLTSSVSRGNVSTPPRHSSGSLTPPTPPITPSSSRSSSTPEPTQI----- 133
DB 428 PQINI-IGSFVQ--NSTVYDSGSHGTAPTLSLQDPDSNNLTPSKABITELSTILE 484
QY 134 -----YRPLRTRNLAPIELHRLTYMSHNSRT--NIKRTFKVDDM-----LSKVE 179
DB 485 ESGSQEFTQFRK---PSHIIQKNPEMPENOLTILNSTSEKWKODDLHTTNAPSIOVD 541
QY 180 KMK-----GEQESHLSA-----HLQLTFTGFPH-KNDKPSPNSE----- 213
DB 542 SKKSEGLIGKQKPAFLSITSCNRSASGYSTDKNEVEFRGFYSARGTKLVGSEALQAK 601
QY 214 -----NEONSV-----TLEVLVVKVCHKKKVDKSCPIRQVPTGKQVP 251
DB 602 KLFSDLENIEETSVEVDFSRSSKYNDVSMVQIEDCNCKNE-----PNKKRL- 652
QY 252 LIPDLNQTGPNPPSLAVSSNEFEPNSHMKVSYLLFR-----VTRPGREFNGM 302
DB 653 -----ILQNIEMTDFVBEYTESYRRNTENEGNQCTDAGRNTCSE 695
QY 303 ING---ETNENIDVNEE---LPARRKRNDEGEKTFVAQMTVFDKNRLQLLDGEYE--- 353
DB 696 SDGSDSKNDTVIHEEENGLPCIDQHNID-----LKLFSQFMKEGNTQIKEGLSDLTC 749
QY 354 VAOQMEECPISKKRATWETILDGKLPPEFTE--SQGPTLQFTLRWTGETNDKSTAPIA 411
DB 750 LVNKABETSHVTMSNKKQJLTANTGQNIKDFDTFYLSFQFASRKNIRSVSESINKARSLLN 809
QY 412 KPIATR-----GESLHQNKPQ-SVKPTQTIAMKESLTTDLQTRKEKDT----- 455
DB 810 QKWTEELNFDSDLSNELLPGLDIKKT-----ISNHEVIENTERKDKITKESLIGENIL 866
QY 456 -----PNNRQKRLFYQFLYNNTRQOTEARDDLHCPCWTLNCRKLYSLKLKHLCHS 509
DB 867 LILQRPESKIKIKESAVLGFHTASGKKIETKE-----SLDKVKNLFEKEQDMS 918
QY 510 REIFNVYHPKGAIDVSNIECYDG-SYAGNPQDIHRQPGFAFRNGPVKRTPIITHILVC 568
DB 919 E-ITNFH--RGAKMSKRECKDDELACGTEITTTTTEYETHSSLEKKVSVSEIAL 975
QY 569 RP-----KRT-----KASMEFELESDGE-----VEQORTYSSGHNH-LYFH 604
DB 976 RPRLSLDNLKYQENLKIASHASQKVDVHENTEKETAKKPTMTVNTQSTYSAIENSPLAFT 1035
QY 605 SDTCLPLRQEMEVDSEDE--KDPFWLREKTIQIIEFSD-----VNEGEKEVWKLWNLHV 658

DB 1036 QDT-----BEKFSVSEASLFEAKWMLRE-----GEWDDQSERINAACKVCKEYDDYV 1083
QY 659 -----MKHGFIADNQMHACHMLFVENYQKIKKGLCRN-FWLHLVSMHD---FNLIS 707
DB 1084 ENPSCGNSNSAITENDKNH-----LSEKQGSTYLSNSTMSNSYHPGFCSHSEVYNKSE 1139
QY 708 IMS-----IDKAVTKLEMQ-----OKLEKGESASP 733
DB 1140 YLSRKIDNSGIBPVIKNIERKNIGPSEIMSPGREADTPQSVNDICVEKLATNSCKN 1199
QY 734 AN-----BEITEQGTANGFSEINSKE-----KALS--TDSVSGVS 768
DB 1200 KNTAKVAISDSNNFNITQKLNDSNNSVPAYSTVNSKRVFAHQTKVTSGETDNCMSVT 1259
QY 769 KQSKKQK 775
DB 1260 KQNTKSK 1266

RESULT 4

US-09-134-001C-5080
Sequence 5080, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynda Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-08-14
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5080
LENGTH: 3696
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 3.8%; Score 156.5; DB 4; Length 3696;
Best Local Similarity 17.4%; Pred. No. 0.00022;
Matches 140; Conservative 147; Mismatches 327; Indels 189; Gaps 32;
QY 30 IEHIEDNHDTPVLEKQELQOPTYVALSYNRFMTDAARREOESLKKIKPQLSLTSL 89
DB 1926 LKINNIPQSTQTKNAQEIINDKAQELQIINN--TPDATEEKEQATNRVNAQLAQAIQ 1984
QY 90 SSVSRGNVSTPPRHSSGSLTPPTPPITPSSSRSSSTPEPTQIYRFLRTRNLIAPIFLH 149
DB 1985 -----NINN--AHSTQEVNESKTSIATIKSVQPNVIKKPTAINSLTQGAN----- 2028
QY 150 RLTLYMGRHSRTNIKK-----TFKVDMLSKVKMKGEQESHLSAH---LQLTFT 199
DB 2029 NQKTLIGNDGNATDDEKAAQQLVTQKLEQIKIHESITQDNQVDNYKAQAITAIKLI-- 2086
QY 200 GFPHKNDKPSNSEQNSVTEVLVVKVCHKKKVDKSCPIRQVPTGKQVPLIPDLNQT 259
DB 2087 -----NANAKRQDAIN---ILTNLAESKSDIRA--NQDATTEENKTAIQSIDDT 2132
QY 260 KPGNFPSLAVSSNEFEPNSHMKVSYLLFRVTRPGRRFNGMTNG----- 305
DB 2133 -----LAQARNNINGANTNALVDENL-----EDGKQLQRIVLSTQTKQAKADIAQA 2180
QY 306 -----ETNENIDVNEELPARKNRE-----DGEKTFVAQMTVFD-KNRLQLLDGE 351
DB 2181 IGQORSTIDQONATTEKQALERLNQETNGVNDRIQAALANQNVTDKRNILLETIRNV 2240
QY 352 YEVAQMEECPISKKRATWETILDGKRLPPFPFSQGPL---QFTLRWTGETNDKSTA 408
DB 2241 EPIVIVPKANEILIRKAAEQTLINQ-----NQDATLEEKQIALGLEEVKNEALN 2292

QY 409 PIKPIATNSBSLHOEN-----KPGSVKPTOTIA--VKESLTDTLOTRKEDTNPENRQK 462
 Db 2293 QVSO--AHNSNDVKIAENNGIAKISEVHPETIIKRNKQEIPODAQS--QIDTINAN---2345
 QY 463 LAIFYQFLYNNTRQOTTEARDLHCPWCTLNCRKLYSLKHLKCHSRFIFYVYHPKGA 522
 Db 2346 -----NKSTNEEKSNAID-----RVNVAKIDAI-----NNTNATT 2377
 QY 523 RIDVINECYDSYAGNPODIHROPGFAPSRNGPVKRTPIITHILVCRPKTKASMESELE 582
 Db 2378 QL---VN---DAKNGSNTSISQILPSTA-----VK-----TNALAALASEAKNNAIIDQ 2421
 QY 583 SEDGEVEQORTYSSGHNRLYFHSIDTCLPLRPOBMEVDEDEKDPWLBEKTIQI-----637
 Db 2422 TPNATAEKEEBAANNKVDRL-----QEEADANILKAHTTDEVNNIKN 2462
 QY 638 BEFSDVNEGEKEVMKMLNHLVHMKGFIADQNMHACMLFVENYQKIIKXLCRNFMHL 697
 Db 2463 QAVQINAVQVEVIKQNVKQNLQFI-DNOKK-----IIENTPDATLEKAEANRLQON 2516
 QY 698 V-----SMHDFNLISIMSIDKAVTKLREMOQKLEKESASPAHEEITSEONGTANGF 749
 Db 2517 VLTSTSDETIANDHNEVDQALDKARPKIEEIVPOVSKRDVLNAIQEAFNSOTQEIQEN 2576
 QY 750 SEINSKEKALETDSVSGVSKQSK 772
 Db 2577 QEATNEEKEALINKINQLLNQAK 2599

RESULT 5

US-07-789-915A-8
 ; Sequence 8, Application US/07789915A
 ; Patent No. 5212058
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Rohan T.
 ; APPLICANT: Tobias, John W.
 ; APPLICANT: Varshavsky, Alexander
 ; TITLE OF INVENTION: Ubiquitin-Specific Proteases
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 ; STREET: Two Militia Drive
 ; CITY: Lexington
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02173
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/789,915A
 ; FILING DATE: 19911108
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brook, David E.
 ; REGISTRATION NUMBER: 22,592
 ; REFERENCE/DOCKET NUMBER: MIT-5091AA
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-861-6240
 ; TELEFAX: 617-861-9540
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 912 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-789-915A-8

Query Match 3.7%; Score 150; DB 1; Length 912;
 Best Local Similarity 18.5%; Pred. No. 9.4e-05;
 Matches 153; Conservative 126; Mismatches 287; Indels 260; Gaps 37;

QY 65 MTDAAARQESLKKTIQPKLSLTSLSSVSRGNVSTPPRHSSGSLTPPVTPPTPSSSFS 124
 Db 3 MQDANKESVMPK-----TSSPPPTPTNMQIPIYQALQVMGYTQ 45
 QY 125 STPTETQI--YRFLTRNLI--APIFLHRT-----LTMSHRN-----STNKKRKT 169
 Db 46 APYLYTQIPAYSF-----NMVQNQPIY-HQSGSPEHLPPONNINGSGSTNNNNKKW 100
 QY 170 KVDMLSKVEKMKGEQSHSLSAHLQLTFTGFHKNDK-----PSPNS-----E 213
 Db 101 HSGIINN-NGSGNGQASGSGSYNKSHTYHNYNNHHPMASPNSSGSGNAGKKOT 159
 QY 214 NEON-----SVTLEVLVVKCHKGRKDVSCPIRQVPTGKQVPLIPDLNQT 259
 Db 160 NSSNGNGSSATSPSYSSYNSQYDLYKFDVTKLN-----LKENSNLTLQLPLF--INTT 213
 QY 260 KPG-----NFPSLAVSNEPEPSN-----SHMVKSYSLLFRVTRGRREFNGM 302
 Db 214 EAFPAASVQVELNKNALNLSLENSVEKSAHHHTKSHS1-----PAHNEE 264
 QY 303 INGETN-ENIDVNEELPARRKRNREDGEXTFVAQMTVFDKNNRRLQLLDGEYEVAMQEMBE 361
 Db 265 VKTETHGESEDAHDKKPHASKDAHELKKKTEVKK-----EDAKQDRNE 307
 QY 362 CPISKKRATWETILDGKRLPPETPSOGTLOFTLRWGTETNDKSTAPIA-----KPLATR 417
 Db 308 KVIQEPQATVLPVVDKKE--PEESVEEN-----TSKTSSPSPPPAKSWSAIASD 356
 QY 418 NGESLHQENKPGSVKPTQTIKVESLTDLQTRKEKDTPENENRQKRIFYQFLYNNNTRQ 477
 Db 357 AIKSRQASNKTVSGSMVTKTPISGITAGVSSINMAAATIGKSSSP1-----LSK 405
 QY 478 QTEARDDLHCPWCTLNCRKLYSLKHLKCHSRFIPNYVYHPKGARIDVINECYDGSYA 537
 Db 406 QPKKDKKYVPSTKGIPLGSA--LRMCFDPDFISYVLRNKDVENKIPVHSII-----458
 QY 538 GNPQDIHROPGFAPSRNGPVKPTPIITHILVCRP-----KRTKASM-----577
 Db 459 --PRGIINRANTCF-----MSSVLQVLLYCKPFDIVNLSTRNTRVGTSSCKLIDA 510
 QY 578 -----SEFLESDG--EVSQORTYSSGHNRLYFHSIDTCLPLRQENVEVDSE 621
 Db 511 CLTMVKQFDKETYEKKFLENADDAEKTETESDAKSKSKSFOHCATADAVKPDF-----565
 QY 622 DEKDPEWLREKTIQIEBFSVDYNEGEKEVMKMLNHLVHMKG-----GFIADQNMHNA 672
 Db 566 -----YKTLSTIPKFDLQWGHQEDAEFLTHLLDQLHHEELISAIDGLTDNE-----612
 QY 673 CMLFVENYQKIIKKNLCNFMHLVSMH---DFNLISIMSIDKAVTKLREMOQKLEGE 729
 Db 613 ----IQNMLQSIINDEQL-KVFFIRNLSRYGKAEP-----IKNASPRKLEIEXY----656
 QY 730 SASPANEETIERONGTANGFSEINSKEKALETDSVSGVSKQSKKQK 775
 Db 657 --GVINDDSTEE-----NGWHE-----VSGSSKRGKTK 683

RESULT 6

US-08-005-002C-8
 ; Sequence 8, Application US/08005002C
 ; Patent No. 5494818
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Rohan T.
 ; APPLICANT: Tobias, John W.
 ; APPLICANT: Varshavsky, Alexander
 ; TITLE OF INVENTION: Ubiquitin-Specific Proteases
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Kevin M. Farrell, P.C.
 ; STREET: P.O. Box 999
 ; CITY: York Harbor
 ; STATE: Maine

```

Qy 578 -----SEPLESDG--EVEQORTYSSGHNRLYFFHSDTCLPLRPQMEVDSE 621
      :|||: : : : : : : : : : : : : : : : : : : : : :
Db 511 CLTWYKQFDKETVEKFELENADDAEKTTESDAKXSSKSKSFQHCATADAVKPDF- 565
      :|||: : : : : : : : : : : : : : : : : : : : : :
Qy 622 DEKDPEWLREKTTITQIEFSDVNGEKEVMKLNHLVMKH-----GFIADNMNHA 672
      :|||: : : : : : : : : : : : : : : : : : : : : :
Db 566 -----YKTLSTIPKFDLQWGHQDAEEFTHLLDQHLHEBLISAIDGLTNE----- 612
      :|||: : : : : : : : : : : : : : : : : : : : : :
Qy 673 CMLEFVENYGQKIKNLCRNFMHLVSMH--DPNLISIMSIDKAVTKLEMQOKLSE 729
      :|||: : : : : : : : : : : : : : : : : : : : : :
Db 613 -----IQMLOSINDEQL-KVFFIRNLSRYGKAEF-----IKNASPLRLKELIEKY 656
      :|||: : : : : : : : : : : : : : : : : : : : : :
Qy 730 SASPANEIETEEQNGTANGFSEINSKEKALETDSVSGVSKQSKKQK 775
      :|||: : : : : : : : : : : : : : : : : : : : : :
Db 657 --GVINDDSTEE-----NGWHE-----VSGSSKRGKTK 683
      :|||: : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-08-487-203A-8
; Sequence 8, Application US/08487203A
; Patent No. 5683904
; GENERAL INFORMATION:
; APPLICANT: Baker, Rohan T.
; APPLICANT: Tobias, John W.
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: Ubiquitin-Specific Proteases
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P. O. Box 999
; CITY: York Harbor
; STATE: Maine
; COUNTRY: U.S.A.
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,203A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/005,002
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: MIT-5091A3Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 207-363-0558
; TELEFAX: 207-363-0528
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 912 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-487-203A-8

Query Match 3.7%; Score 150; DB 1; Length 912;
Best Local Similarity 18.5%; Pred.No.9.4e-05;
Matches 153; Conservative 126; Mismatches 287; Indels 260; Gaps 37;

Qy 65 MTDAAREQESLKKIQKPLSLTLSSVSRSRGNVSTPPRHSSGSLTPFPVTPITPSSSPRS 124
      :|||: : : : : : : : : : : : : : : : : : : : : :
Db 3 QMDANKEESVMYPK-----TSSPPPTTNQIILYQAPLQMYGYTQ 45
      :|||: : : : : : : : : : : : : : : : : : : : : :

Qy 125 STPTPEPQI--YRPLRTRNLI--APFLHRT-----LTMVSHRN-----SRTNKRTFF 169
      :|||: : : : : : : : : : : : : : : : : : : : : :
Db 46 APLYLPTQIYAYSF---NMVQNQPIY-HQSGSPHPLFPQNNINGSGTTNNNNINKKW 100
      :|||: : : : : : : : : : : : : : : : : : : : : :

```

```

QY 170 KVDDMLSKVKKQSGSHLSAHLQLTFTGFHKNDK-----PSPNS-----E 213
Db 101 HNGTINN-NGSSNGQAGNSGSGMYNKSHTYHNNHIPPMSNSGNSNMGKQOT 159
QY 214 NQCN-----SVTLEVLVVKVCHKRDVSCPIRQVPTGKKQVPLIPDLNQ 259
Db 160 NSSNGSGSATSPSYSSNSQYDLKFDVTKLN-----LKENSNLILQLPLF--INTT 213
QY 260 KPG-----NFPSSLAVSSNEFPSP-----SHMVKSYLLPVTPRGRREFNGM 302
Db 214 EAEFAAASVQVELNWKALNLSSELSSESVSEKSAHHHTKSHSI-----PKNEE 264
QY 303 INGETN-ENIDVNEELPARRKENREDGKTFVAQMTVFDKNNRLQLLDGEYVAMQEMEE 361
Db 265 VKTETHGEEDAHKKPHASKDAHELKKKTEVKK-----EDAKQDENE 307
QY 362 CPISKKRATWETILDGKRLPPPTFSQGTLOFTLRWGTENDKSTAPIA-----KPLATR 417
Db 308 KVIQEQATVLPVVDKKE--PEESVEEN-----TSKTSSPSPSPAASWASATSD 356
QY 418 NSESLHQENKPGSVKPTQTIIVAKESLTTDLOTRKEKDTPENENROKLRIFYQFLYNNNTRO 477
Db 357 AIKSRQASKTVSGMWTKTPISGTTAGVSSNNMAATIGKSSPL-----LSK 405
QY 478 QTEARDLLHCPWCTLNCRKLSLLKHLKCHSRFIFNVYHPKGAIDVSNICYDGSYA 537
Db 406 QPKKDKKYVPPSTKGIPLGSLA--LRMCFDPDPISYVLNRKOVENKIPVHSII-----458
QY 538 GNPQDIHQPGFAPSGFVKEPTPIHILVCEP-----KRTKASM-----577
Db 459 --PRGILIRANICF-----MSVLQVLLYKPFIDVINLSTNTRNSRVGTSSCKLLDA 510
QY 578 -----SEFLESDG--EVEQRTYSSGHNRLYFHSDTCLPLRQPEMEVDSE 621
Db 511 CLTMVKQFDKETYKKFLENADDAKTTESDAKSKSKSFQHCATADAVKPDF-----565
QY 622 DKQPEWLREKTIQIEFSDVNEGEKVKMLNHLVWKH-----GFIADQNQVHA 672
Db 566 -----YKTLSTIPKFKDQWQHQBDAEFLLDQJHEBLISAIDGLTONE-----612
QY 673 CMLFVNYGQKIIKGNLORNFMLHJVSMMH--DFNLISIMSIDKAVTKLREMOQKLEKGE 729
Db 613 ---IQNMQSLNDQL-KVFFIRNLSRYGKAEP-----IKNASPLKELIEKY----656
QY 730 SASPANEITEEQNTANGFBEINSKEXALETDSVSGVSKSKQKQK 775
Db 657 --GVINDDSTEE-----NGWHE-----VSGSKRGKRTK 683

```

RESULT 8

```

US-08-118-101A-2
; Sequence 2, Application US/08118101A
; Patent No. 5620892
; GENERAL INFORMATION:
; APPLICANT: Kurtz, Stephen E.
; APPLICANT: Knickerbocker, Aron M.
; APPLICANT: McCullough, John R.
; TITLE OF INVENTION: A STRAIN OF SACCAROMYCES CEREVISIAE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/118,101A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; REFERENCE/DOCKET NUMBER: DC27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-5901
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-118-101A-2
;
; Query Match 3.6%; Score 146.5; DB 1; Length 1235;
; Best Local Similarity 20.1%; Pred. No. 0.00033;
; Matches 154; Conservative 90; Mismatches 227; Indels 295; Gaps 39;
;
QY 134 YRFLRTPNLIAPFLHRTLTYSMSHRSRTNIRKTKFKVDDMLSKV-----EWMK 182
Db 148 FKMRRTKTLERELTARTMTKRTGTQRTSYPRKQAKTDDFOEKLFGEMVNRDEQDSVH 207
QY 183 GEQESHLSAHLQLTFTGF-----FHKNDKPSNSENENQNSVTLVLLVKVCHKKR 233
Db 208 SDQSHDIDSRDNNNTNHNHSGSSGLDDFVKEDETTDNGEYQENN-----252
QY 234 KQVSCPIRQVPTGKKQVPLIPDLNQTKGNFPLAVSSNEF--EPNSNMVKSYSLFP--289
Db 253 -----SYSTVGSSNTVADESINOKPSPSLRFPDEP 283
QY 290 -RVTPRGRREFNGINGETNENIDVNEELPARRKENREDGKTFVAQMTVFDKNNRLQL 348
Db 284 HSKQRPAP-----VPSEKFAKRGSRD-----ISPADMYASIMWL 318
QY 349 DGEYVAMQEMEECP-----SKRATWETILDGKRLPPPTFSQGTPL 392
Db 319 QQKHE-ATADEGPPLVIGSPADGTRYKSNVNLKAKTG--INGNKIKIRDKNESNTD 374
QY 393 QFTLRWGTNDKSTAPIAKPLATNSLSLHCE--NKPGSVKPTOT-----IAVKESL 443
Db 375 QNSV--SSEAN--STASVS-----DESSLHTNFGNKVPSLR-TNTHRSNGPIAITDNA 423
QY 444 TTDLOTRKEKDTPENENROKLRIFYQFLYNNNTR--QOTEARDDLLHCPWCTLNCRKLYS-499
Db 424 ETD-----KKGPS-----IQFDITKPPRKISKVSTFDDLN-FKSSVLYRKKASK 468
QY 500 --LLKHLKLCCHSRFIFNVYHPKGAIDVSI-NECYDGSYAGNPODIHQPGFAPSRNGP 556
Db 469 KYLMKHF-----PKASRIIRQIKRRLSTGSIENK-----SSNNV 502
QY 557 VKRTPITHILVCRPKRTKASME-----FLESEDEGEVQORTYSSGHNRL 601
Db 503 SQRKPT-----DMDDDDDDDDGDNNDNEEYFADNESGD-EDERVQOSEP---546
QY 602 YFHSOTCLPLRQPEME-----VSEDEKDEPEWLREKTIQIEFSDVNE 645
Db 547 --HSDSELKSHQOOQEKHQQLQNLHRMYKTKSFDDNRGRAVPMESRSTI-DMAEAKDNE 603
QY 646 -----GEKEVMKLNHL-----VMKHG-FIADQNQNHACMLFVENYQOK 683
Db 604 LARTPDFQWYQNNWKAHKKKPNFRKGWNNKIFEHGFYASDSDRNYPDN---SNTG--658
QY 684 IIKKNLCRNFMHLVSMHDFNLISIMSIDKAVTKLREMOQKLEKESASPANEITEEQN 743
Db 659 -----NSILHYAE-----SILHEDGS-----HKNGBEASDSNENIYSTNG 695
QY 744 GT-----ANGFSEINSKEKA-----LETDSVSGVSKSKQKQK 776
Db 696 GSHNGNLNPTYNDDDEGYGLHFDTDYDLDPRHDLKSGSGKTYL 741

```

```

RESULT 9
US-09-418-710-21
; Sequence 21, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-21

Query Match 3.5%; Score 142.5; DB 4; Length 1972;
Best Local Similarity 17.4%; Pred. No. 0.0017;
Matches 146; Conservative 131; Mismatches 301; Indels 263; Gaps 33;

QY 30 IHHEDNHDTPRVLKQELQOPTYVALSYINRMTDAARQESLKKKIQPKLSLTL 89
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
83 IESEDDSDSEAEQHSNNQ----VLLHGISDPKADQKATEKAQEKRIHQPLAFE 138
QY 90 SSVSRGNVSTPRHS--SGSLTPV---TPITPSSFRSSTPTPTQIVRFLRNL 143
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 SGT-----HFSQSQKQPVLSQQLPFIQSSQAKESVNHKHSVIGSLGVNV 188
QY 144 APIFL---HRTLYMS-----HRNRTNKRKTKVDMDLSKVE-----RMKG 183
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 KPLSLVNAQAKETYMKLI VPSDVLKAGNKTSESSLTSELRSKREYKQAPPSQLK 248
QY 184 EESHSLS-----AHLQLTFTGFHKNDKSPNSN-----EQNSYTL 221
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249 QESSKSLKVI AALNPKATSSPAHPQTL-----ENNHPNPFITNALLGNHQPNG 303
QY 222 EVL-----LVVYCHKRKXDV-----SCPIRQVPTGK---QVPLIPDLNQT 260
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 SVIQEAPLALTTKTKWQSKINENIAAASSTPSSPVNLSTSGRTPGNTQTPWPS 363
QY 261 PGNFSLAVSNPEP-----SNHVKYSYLLFRVTR---PGRREFNGMNGEIN 313
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 HSQGEKAVSNN--VNPVKTQHSHPAKSLVEQFRGTDSDIPSKD-----SE 417
QY 314 NEELPARRKRNDEKTFVAQMTVFDKMRRLQLDGEYVAMQWESCPISKRAWET 373
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 DDE-----EEDDEDE-----DDESDDSDSDSDSDSDSDSDSDSDSD 453
QY 374 ILDGKRLPFFTFSGPTLQFLRWGTGENDKSTAPIAKPLATRNSES----- 429
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 DDDKDDQDESDDSDTEGKTSML-----NKTSSVKSPSMLTGHSTPRNLHIAK 506
QY 430 SV-----KPTQIYAVKESLTTDQTRKEKDTFNENR---QKLRIFYQFLYNN 476
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507 SAPAALCESQSPAFGLTSSSTLTSPPHSGTSKRRRVTDRELRLPLEYGMQRETR 566
QY 477 -QOTEARDLHCPWCTLNCRLKLYSLKHLKCHSRFIFNYVHPKGAIDVSI 535
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 GGRLQGEVAYAP-CGKKLRQYPEVILKYL-----RNGIMDIS----- 603
QY 536 YAGNFQDTHRQGPAPFNGPVKRTPIHILVCRPRTKASMSFLESEDEGEQRT 595
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
604 -----RDNFSFS-----AKIRVGDFYEARDGPQEMQWC-- 631

Query Match 3.4%; Score 137.5; DB 4; Length 2354;
Best Local Similarity 19.7%; Pred. No. 0.0094;
Matches 162; Conservative 140; Mismatches 300; Indels 221; Gaps 41;

QY 21 LHPTADLLEHEDNHDTPRVLK-----KQELQOPTYV-----ALSINRPM 69
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1476 LETPLOAD---HSEDS-IDRRSEEMIKVLGEXLERNQYLLERLQBEKLSNKL-- 1529
QY 70 RREGES---LKKXIQPKLSLTLSSS-VSRGNVSTPRHSSG-----S 107
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1530 QKEMETSVLLKDDLQOKLESLLSENILKENIDITLKHSDTQAOQKTOQLAKNLA 1589
QY 108 LTPPEVTPPTPSSFRSSTPTPTQIVRFLRNLIAPIFLHRTLYMSHRNRTNKR 167
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; TITLE OF INVENTION: Chromosome Congression
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 1
; LENGTH: 2354
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
; OTHER INFORMATION: tail domain
US-09-150-867-1

Query Match 3.4%; Score 137.5; DB 4; Length 2354;
Best Local Similarity 19.7%; Pred. No. 0.0094;
Matches 162; Conservative 140; Mismatches 300; Indels 221; Gaps 41;

QY 21 LHPTADLLEHEDNHDTPRVLK-----KQELQOPTYV-----ALSINRPM 69
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1476 LETPLOAD---HSEDS-IDRRSEEMIKVLGEXLERNQYLLERLQBEKLSNKL-- 1529
QY 70 RREGES---LKKXIQPKLSLTLSSS-VSRGNVSTPRHSSG-----S 107
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1530 QKEMETSVLLKDDLQOKLESLLSENILKENIDITLKHSDTQAOQKTOQLAKNLA 1589
QY 108 LTPPEVTPPTPSSFRSSTPTPTQIVRFLRNLIAPIFLHRTLYMSHRNRTNKR 167
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 1590 IAASDNCPIQKE-TSADCVHPLEKILLTEEL-----HQXTNQEOK 1632
QY 169 TFKVDDMLSKVB-KMKQGEHSLSAHLQLTFTFFHNDKPSNS-ENEONSVTLEVL 225
Db 1633 LLHEKNELEQAQVELKCEVE-HLMKSMIE-----SKSLESQHEKHDEQOILLA 1691
QY 226 VVCHKGRKDVSCPIRQVPTGKQVPLIPDLNQTGPGNFPPLAVSSNEFEPSNHHVKS 285
Db 1682 LK-----QMQVVTQEK-----ELQOTHEHLTAEV 1707
QY 286 SLFVTRPGREFNGMETNENIDVNEELPARKNR-----EDGEKTFV 333
Db 1708 DHLKENIELGLNFKNEAQKTKTEQCLLNENKELEQSOHRLQCEIEELMKSLDKES--- 1764
QY 334 AQMTVPDKNRRLQLLDGVEYAMQMECPISKKRATWETILDGKRPUPFTFGQGTLO 393
Db 1765 ALETKESEQVINUNQENEMVMEELKNSOR-----TVI-----AERDOLQ 1808
QY 394 FTLRWGTNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLTTDLQTRKEK 453
Db 1809 DDLRESSVMSIETODDLKA-----QEAQQQKOK-----VQELTSQISVLQEK 1852
QY 454 DTPNENROKLRIFYOFLYNNTRQOT-BARDDLHCPWCTLNCRKLYSLLLKHLKLC--HSR 510
Db 1853 ISLLEN-----QMLYNVATVKETLSERDOLN-----OSKQHLFSEIETLSLSLKE 1899
QY 511 FIFNVVYHPK-GARIDVINECYDGSYAGNPQDIHQPGFAGSRNGPVKPTPTH--IL 566
Db 1900 FALQAEKDKADAAKTIDITE-----KISNIEEQ-----LLQOATNLKLTLYERSLI 1948
QY 567 VCRPK-RTKASMSFLESSED--GEVQOQTYSSGHNRLLYFHSDDTCLPLRPQEME----- 617
Db 1949 QCKEALANTEHLRETLKSKOLALGMEQERDEAA--NKVIALTEKMSLEEQINENVT 2006
QY 618 -VDSDEKDPWL---REKTIQIEEPFD-----VNEGEKEVM-----KWLNLHVMK 660
Db 2007 LKEGEKETYLPQPSKQSSQVEELRESLKYDQLEAEKEISEATNEIKNL-TAK 2065
QY 661 HGFIADNQMNHACMLFVNYGQKIKKNCNFMHLVSMHDFNLISIMSIDKAVTKLRE 720
Db 2066 ISSLEETLQNASIL-----NEAVSERENLRHSQQLVSELEQLSLTLKSRDHAPA--- 2116
QY 721 MOOKLEKESASANEETIEQNGTANGFINSKEKALETDS 763
Db 2117 -QSKREKDEAVNKI--ASLAEBIKILTKEMDFRDSKESLQEQS 2157

RESULT 11

US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET B.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 3.3%; Score 135; DB 1; Length 3248;
Best Local Similarity 20.3%; Pred No. 0.019;
Matches 180; Conservative 134; Mismatches 339; Indels 232; Gaps 44;

QY 21 LHPFTLADLI-----BH--IEDNHIDTDPVLEKQELQOQPTTV-ALSYNRMWTDAAARE 72
Db 1402 MHFAELQEKFLSQSEHKILHQCMSKMSLQ-----TYVDSLKAENLVSLTNLENF 1456
QY 73 QBSLKKIKPKLSLTLSSSVSRGNVSTPRHSSGSLTPVTPPTPSSSFRSSTTEPTQ 132
Db 1457 QGDLVKEMQUGL-----EEGLVPSLSSCVDPSSSLSS--LGDS 1494
QY 133 IYRFLRNLRIAPIFILHRTLYTHSRNRTNKRITFKVDDML-----SKVKMKGEOS 187
Db 1495 FYRAL-----LEQTGMSLLNLEGAVSANQCSVDEVFCSSLQTYVDSLKA--EN 1542
QY 188 HSLSAHLQLTFTGPHKNDK-----PSPNSENENQSVILEVL-----LVKVCHKRD 235
Db 1543 LVLSTNLNLF-NFQGDVLEKQGLGLEGLVPSLSSCVDPSSSLSSLGDSFYRALLEQTGD 1601
QY 236 VS-----CPIROV-----PTGKQVPLIP-----DLNQTGPNPPLAV 269
Db 1602 MSLLNLEGVSNQCSVDEVFCSSLQENLTKETPSAPAKGVEELESCEVYRQSLK 1661
QY 270 SNEFEPSNHHVKSYSLLFRVTRPGRRFNGMNGETNEN-----IDVNEBLP 318
Db 1662 LEEKVESQGIKMKKEIQELEQLLSERQELDLCKRQYLSENEQMOQKLTSTVLEMSKLA 1721
QY 319 ARKRNRDEGKTFVA--QMTVFDKNRRLQLLDGVEYVAMO--EMECPISKKRATWETIL 375
Db 1722 AEKQTEQLSLEVARLQGLDLSR--SLLGITEDAIQGRNESCDSIHEHSEIT-- 1778
QY 376 DGKRLP-----PFETFSQGTPLQFTLRMTGETN-DKSTAPIAKPLAT 416
Db 1779 --ERTPKHDVHOICDKDAQDQDLNLDIEKITETGAVKPTGECGSGSPDTPNPPGDKTQ 1836
QY 417 RNESLSHQ--ENKPGSVKPTQTIIVKESLTTDLQTRKEKDTENENROKLRIFYQLYNN 474
Db 1837 GSSECISELFSFGPNALVPMDFLGNQEDI-HNLQLR-VKETSNEMLRLHVI----- 1886
QY 475 TRQOTEARDLHCPWCTLNCRKLYSLLLKHLKLCNHSRFFIFNYVYHPKGIARDVINECYD- 533
Db 1887 -----EDRD-----RKVESLLNEMKELDSKL-----HLQEVQLMTKEACIEL 1924
QY 534 ----GSYAGNPQDI-HRQPGFAGSRNGPVKPTPTHIL-----VCRPKTKASKE----- 579
Db 1925 EKIVGELKENSDDLSEKLEYFSCDHQELLQORVETSEGLNSDLEMHADKSSREDIGDNVAK 1984
QY 580 -----FLESEDEGEVQQRYSYG--HNRLYFHSDDTCLPLRPOEMEVDSE-----EKD 625
Db 1985 VNSDKERFLDVEN-ELSRIRSEKASIEHEALYLEAD-----LEVVQTEKLCLEKD 2034
QY 626 PEWLREKTIQIEEFSDVNEGEKEVKNLWNLHVMKHGFIADNQMNHACMLFVNYGQKII 685


```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254-6

Query Match          3.2%; Score 132; DB 1; Length 2482;
Best Local Similarity 20.1%; Pred. No. 0.023;
Matches 170; Conservative 125; Mismatches 301; Indels 250; Gaps 43;

QY 21 LHFPTLADLI-----EH--IEDNHIDTPRVLEKQELQOPTYV-ALSYINRFMTDAARRE 72
DB 770 MHFAELQBFJLSQSEHKILHDQCMSSKVSQELQ-----TYVDSLKAENLVLSTNLRNF 824
QY 73 QESLKKIKQPKLSLTLSSSVSRGNVSTPPRHSSGSLTPPVTPPTPSSSFSSFTPTPTQ 132
DB 825 QGDLVKEQGL-----ESGLVPSLSSCVDPSSSLSS--LGDSS 862
QY 133 IYRFLRTRNLAPIFLHRTLYTMSHRNRTNIKRTFKVDDM-----LSKVE----- 179
DB 863 FYRAL-----LEQTGMSLLSLNLAGAVSANQCSDVEVFCSSLQSEENLTKETPSA 912
QY 180 KMKGEQSHSLSAHLQLTFTGFFHKNDKPS--PNSENEQNSVTLEVLVVKVCHKKRQDVS 237
DB 913 PAKGVVEELSECEVYRQSLKLEKEMESQGIWKKEIQE-----LEQLL-----SSERQELD 964
QY 238 CPIRQVPTGKKQVPLIPDLNQTGPNFPPLAVSSNEPEPSSNHWKSVSYLLFRVTRPGR 297
DB 965 CLRQKY-----LSENE----- 975
QY 298 EFNGMINGETNENIDVNEELPARKKRNREDEKTFVA--QMTVDEKRRRLQLLDGEYVA 355
DB 976 QWQQL-----TSVTLEMSKLAABKQTEQSLSELEVARLQGLDLSR--SLGIDTEDA 1031
QY 356 MQ-EMEECPISKGRATWETILDGKRLP-----PPETFSQOPTLQFTL 396
DB 1032 IQGRNESCDSKEHTSEIT-----ERTPKHDVHQICDKDAQDNLNDIEKITEGALKPTG 1087
QY 397 RWTGETN-DKSTAPIAKPLATRNESLHQ--ENPGSVKPTQTIAVRESLTTDLQTRKEK 453
DB 1088 ECSGEQSPDTNYEPGDKTGSGECISSELSFGPNALVPMDFLGNQEDI--HNQLRL-VK 1145
QY 454 DTPENRQKLRIFYQLYNNNTRQQTARDDLHCPWCTLNCRLKLYSLKELKCHRSFIF 513
DB 1146 ETSNENLRLVLI-----EDRD-----RKVESLLNEMKELDSKL-- 1179
QY 514 NYVHPKGAIRDVSNIECYD-----GSYAGNPQDI-HRQGFAPSRNGPVKRTPIITHIL- 566
DB 1180 ----HLQEVQLMTKIEACIELEKIVGLKKNESLSEKLEVFYSCDHQELLQRTVETSEGLN 1235
QY 567 ----VCRPKRTKASMS-----FLESDEGEVQOQRTYSSG--HNRLYFHSDDC 608
DB 1236 SDLEWHADKSRREDIGDNVAKVNDWSKRFDLVDEN--ELSRSEKASIEHEALYLEAD-- 1292
QY 609 LPLRPQMEVDSDE-----EKDPEWLREKTIITQIEFSDVNEGEKVMKLNWLVHMKHGF 664
DB 1293 ----LEVQTEKLCLEKDNE-NKQKVIYCLBEELSVMVTSERNQLR-GELODTMSKKT 1343
QY 665 ADNQMHACMLFVENVQKILIKKLNCRNFMHLVSMHD-----FNLISIMSD-----K 713

```

```

DB 1344 ALDQLS-----EKKMKEKTQELSHQSECLHICQVAAEAVKEKTELLQTLSSDVSSELLK 1396
QY 714 AVTKLEMOQKLEKGESAPANEETEEBQNGTANGFSEINSKEKAL----ETDSVSGVSKQ 770
DB 1397 DKTHLOEKQSLKESQALSUTKCELENO-----IAQLN-KEKELLVKESSELQARLSE 1449
QY 771 SKQOKL 776
DB 1450 SDYEKL 1455

RESULT 14
US-08-938-830-26
; Sequence 26, Application US/08938830
; Patent No. 6040437
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; APPLICANT: Dowbenko, Donald J.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPs)
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,830
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 07-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1066P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-938-830-26

Query Match          3.2%; Score 128.5; DB 3; Length 907;
Best Local Similarity 19.2%; Pred. No. 0.0096;
Matches 98; Conservative 74; Mismatches 174; Indels 165; Gaps 22;

QY 18 GGLHPTLADLIEHIEDNHIDTPRVLEKQELQOPTYVVALSYINRFMTDAARREQESLK 77
DB 262 GTGTITPPLPEFNDYFKENGLNYD---IDQLISKAPSY---FYSSSRPSASASSPTR 315
QY 78 KKIQPKLSLTLSVSRGNVSTPPRHSSGSLTPPVTPPTPSS----- 120
DB 316 SAFRPKTSETVSSEV-----VSSPP-----TSPHSPVKPVSNVEQVEQVTEVELSIPVP 364
QY 121 SFRSTPTPTPTQIYRFLRTRNLIAPIF--LHRTTYTMSHRNRTNIKRTFKVDDMLSKV 178
DB 365 STQEAESQKPLVTGSSMRPSPVTSPTFEVAARPLTSMDSRSSH----- 407
QY 179 EWMKEQESHSLSAHLQLTFTGFFHKNDKPSP---NSENEQNSVTLEVLVVKVCHKKRQD 235
DB 408 ---NAETEVOQIPAATDI-----SPEVKEGKNSENAIT-----KONDD 442

```


QY 236 V--SCPIRQVTKGKQVPLIPDL-----NOTKPGNPPSLAVSSNEFPSPNSHMVKSYSLL- 288
Db 443 IILSSQLQPTATGSRSSRLSFRHGHGSQTSILGSIKRKSIMRMGRPTSPFPMGSSFSNMG 502
QY 289 FRVTRPGRRREF-----NGMINGETNENIDVNE-----ELPARRKRNRD 327
Db 503 SRSTSPTEKGFASQHQHATGASVQSDLEDDIDPRANVLVGNPNMLSVGEAPVESTSKEED 562
QY 328 GE-----KTFVAQMTVFDKRRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPP 382
Db 563 KDVPDPIANAMAEELSSSMRRRQSTSV-----DEAPVSLSKTSSSTRLNG----- 607
QY 383 PETFSQPTLOFTLRWTGTEND-----KSTAPIAKPLATRNSESLHGOENKPGSVKPTQT 436
Db 608 -----LGVHSRNTSIASDIDGVPKKSTLG-APPAHTSAQMORMNSFAS----- 651
QY 437 IAVKESLTTDLOTRK---EKDTPNENRQKLR 464
Db 652 -----QTKQVFGQRTENSARESLR 671

RESULT 15
US-09-020-222-26
; Sequence 26, Application US/09020222
; Patent No. 6111073
; GENERAL INFORMATION:
; APPLICANT: Lasky, Laurence A.
; TITLE OF INVENTION: Tyrosine Phosphorylated Cleavage
; TITLE OF INVENTION: Furrow-Associated Proteins (PSTPIPS)
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,222
; FILING DATE: 06-Feb-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/798419
; FILING DATE: 02/07/1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1066r1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-09-020-222-26

Query Match 3.2%; Score 128.5; DB 3; Length 907;
Best Local Similarity 19.2%; Pred. No. 0.0096;
Matches 98; Conservative 74; Mismatches 174; Indels 165; Gaps 22;
QY 18 GCGLHPTLADLIEHDNHDTPRVLEKQLQOPTVALSYINRFMTDAARQESLK 77
Db 262 GTGTIPPLPEPDFKENGKLYD---IDQLISKAPSY---PYSSRPSASASLASSPTR 315
QY 78 KKIQPKLSLTLSSSVSRGNVTPPRHSGSLTPPTPTPSS----- 120

Db 316 SAFRPKSTETVSSEV-----VSSFP-----TSPLHSPVKPVSNQEQVTEVELSIPVP 364
QY 121 SFRSSTPTPTQIYRFLRTRNLIAPIF--LHRTLITVMSHRNSETNIRKRTFKVDDMLSKV 178
Db 365 SIQEAESQKPEVLGTSSMRSPSVTPTFEVAARPLTSMVRSRH----- 407
QY 179 EKKMGQESGSHLSAHLQLTFTGFFHKNDKPSP---NSENEQNSVTLEVLVVKVCHKKRD 235
Db 408 ---NAETEVOAIPAATDI-----SPEVKEGKNSENAIT-----KDND 442
QY 236 V--SCPIRQVTKGKQVPLIPDL-----NOTKPGNPPSLAVSSNEFPSPNSHMVKSYSLL- 288
Db 443 IILSSQLQPTATGSRSSRLSFRHGHGSQTSILGSIKRKSIMRMGRPTSPFPMGSSFSNMG 502
QY 289 FRVTRPGRRREF-----NGMINGETNENIDVNE-----ELPARRKRNRD 327
Db 503 SRSTSPTEKGFASQHQHATGASVQSDLEDDIDPRANVLVGNPNMLSVGEAPVESTSKEED 562
QY 328 GE-----KTFVAQMTVFDKRRRLQLLDGEYEVAMQEMEECPISKKRATWETILDGKRLPP 382
Db 563 KDVPDPIANAMAEELSSSMRRRQSTSV-----DEAPVSLSKTSSSTRLNG----- 607
QY 383 PETFSQPTLOFTLRWTGTEND-----KSTAPIAKPLATRNSESLHGOENKPGSVKPTQT 436
Db 608 -----LGVHSRNTSIASDIDGVPKKSTLG-APPAHTSAQMORMNSFAS----- 651
QY 437 IAVKESLTTDLOTRK---EKDTPNENRQKLR 464
Db 652 -----QTKQVFGQRTENSARESLR 671

Search completed: August 25, 2004, 17:58:57
Job time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 25, 2004, 18:15:39 ; Search time 407.5 Seconds
(without alignments)
6114.120 Million cell updates/sec

Title: US-09-874-162A-7
Perfect score: 7690
Sequence: 1 cccgcgcgcgcgcgcag.....tacataaagtccaataatat 4409

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlp

-Q/cgn2_1/USPTO.spool_p/US09874162/runat_25082004_171843_2162/app_query.fasta_1.4551
-DB=A_Geneseq_29Jan04 -OFFT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09874162 @CGN1.1.642 @runat_25082004_171843_2162 -NCFU=6 -ICFU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMECUT=120 -WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|--------|-------|--------|------------|--------------------|
| 1 | 3725 | 48.4 | 776 | 5 ABG30843 | ABG30843 Human jAZ |
| 2 | 3413 | 44.4 | 803 | 3 AAB00066 | AAB00066 KIAA0160 |
| 3 | 3237 | 42.1 | 739 | 5 ABG30842 | ABG30842 Human joi |
| 4 | 2043 | 26.6 | 388 | 4 AAU15978 | AAU15978 Human nov |
| 5 | 2043 | 26.6 | 388 | 6 AAU15947 | AAU15947 Human nov |
| 6 | 1497 | 19.5 | 289 | 4 AAU15958 | AAU15958 Human nov |
| 7 | 1497 | 19.5 | 289 | 6 ABUS5027 | ABUS5027 Human nov |
| 8 | 1453.5 | 18.9 | 292 | 4 AAU16416 | AAU16416 Human nov |
| 9 | 1453.5 | 18.9 | 292 | 6 ABUS5485 | ABUS5485 Human nov |
| 10 | 1127.5 | 14.7 | 955 | 4 ABB71635 | ABB71635 Drosophil |

| | | | | | |
|----|-------|------|------|-------------|---------------------|
| 11 | 940 | 12.2 | 175 | 4 AAU16402 | AAU16402 Human nov |
| 12 | 940 | 12.2 | 175 | 6 ABUS5471 | ABUS5471 Human nov |
| 13 | 733 | 9.5 | 278 | 4 AAU16381 | AAU16381 Human nov |
| 14 | 733 | 9.5 | 278 | 6 ABUS5450 | ABUS5450 Human nov |
| 15 | 661 | 8.6 | 243 | 7 ADB64250 | ADB64250 Human pro |
| 16 | 661 | 8.6 | 243 | 7 ADC06802 | ADC06802 Human pro |
| 17 | 564 | 7.3 | 143 | 4 AAU15931 | AAU15931 Human nov |
| 18 | 564 | 7.3 | 143 | 6 ABUS5000 | ABUS5000 Human nov |
| 19 | 490 | 6.4 | 243 | 5 ABG30841 | ABG30841 Human jux |
| 20 | 457 | 5.9 | 102 | 3 AAG03182 | AAG03182 Human sec |
| 21 | 429 | 5.6 | 202 | 3 AAB43387 | AAB43387 Human ORF |
| 22 | 280 | 3.6 | 171 | 4 ABB11690 | ABB11690 Human SFP |
| 23 | 238.5 | 3.1 | 604 | 3 AAY53933 | AAY53933 An Os-MPC |
| 24 | 232 | 3.0 | 241 | 4 ABG26463 | ABG26463 Novel hum |
| 25 | 230 | 3.0 | 498 | 3 AAG54168 | AAG54168 Arabidops |
| 26 | 227.5 | 3.0 | 367 | 3 AAG54169 | AAG54169 Arabidops |
| 27 | 227 | 3.0 | 767 | 4 ABB58240 | ABB58240 Drosophil |
| 28 | 224 | 2.9 | 611 | 3 AAY53932 | AAY53932 A MPC1 pr |
| 29 | 221 | 2.9 | 445 | 3 AAB00060 | AAB00060 VRN2 poly |
| 30 | 217.5 | 2.8 | 440 | 3 AAB00061 | AAB00061 VRN2 poly |
| 31 | 212 | 2.8 | 241 | 4 ABG26463 | ABG26463 Novel hum |
| 32 | 211 | 2.7 | 692 | 6 ABP70524 | ABP70524 Histone d |
| 33 | 211 | 2.7 | 813 | 3 AAB01674 | AAB01674 FIS2 prot |
| 34 | 188 | 2.4 | 2013 | 3 AAB18265 | AAB18265 Plasmodiu |
| 35 | 176.5 | 2.3 | 1142 | 4 AAG70713 | AAG70713 S cerevis |
| 36 | 176.5 | 2.3 | 1142 | 6 ABR53123 | ABR53123 Protein s |
| 37 | 176.5 | 2.3 | 1349 | 4 AAG70854 | AAG70854 C albica |
| 38 | 161.5 | 2.1 | 583 | 5 ABG93227 | ABG93227 C. albica |
| 39 | 161 | 2.1 | 1875 | 6 ABR53560 | ABR53560 Protein s |
| 40 | 159.5 | 2.1 | 339 | 3 AAG54170 | AAG54170 Arabidops |
| 41 | 159.5 | 2.1 | 912 | 2 AAB36731 | AAB36731 Ubiquitin |
| 42 | 157.5 | 2.0 | 1388 | 7 ADB80468 | ADB80468 Ovarian c |
| 43 | 156.5 | 2.0 | 3696 | 5 ABA040235 | ABA040235 Scaphyloc |
| 44 | 155 | 2.0 | 1087 | 2 AAY19935 | AAY19935 B. burgdo |
| 45 | 155 | 2.0 | 1119 | 2 AAY19934 | AAY19934 B. burgdo |

ALIGNMENTS

RESULT 1
ABG30843
ID ABG30843 standard; protein; 776 AA.
XX AC ABG30843;
XX DT 21-OCT-2002 (first entry)
XX DE Human jAZF1/jJAZ1 fusion protein.
XX KW Human; jAZF1; juxtaposed with another zinc finger; jJAZ1; jAZF1/jJAZ1;
XX KW joined with jAZF1; proliferation; endometrial stroma tumour; immunogen;
XX KW antigen; antibody; fertility; pregnancy; gene therapy; vaccine;
XX KW fusion protein.
XX OS Homo sapiens.
XX OS Synthetic.
XX OS Chimeric.

Key Location/Qualifiers
FT Misc-difference 13..37
FT /note= "Encoded by ACCTGCGATTCGGGGCTCGGACTCCTTC-
FT CCACCTGCGGACCTCATCGACATCGAGCACACAC. C2H2 zinc
FT finger domain"
FT Region 99..100
FT /note= "Breakpoint for production of fusion protein
FT jAZF1/jJAZ1"
FT Misc-difference 357
FT /note= "Encoded by CAG"
FT Misc-difference 485..508
FT /note= "Encoded by CTGATTGCGCTGGTGTACTCTGAATCGCGAA-
FT ACTTTATAGTTTACTCAAGCATCTTAACCTCGCCAT. C2H2 zinc finger
FT domain"
FT Region 558..575

/note= "Bipartite nuclear localisation signal"

FT XX
PN XX
XX XX
PD 13-DEC-2001.
XX XX
PF 04-JUN-2001; 2001WO-US017936.
XX XX
PR 02-JUN-2000; 2000US-0209093P.
XX XX
PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX XX
PI Koontz J, Sklar J;
XX XX
DR WPI; 2002-575047/61.
DR N-PSDB; ABK89163.
XX XX
PT Novel JAZF1, jJAZ1 or JAZF1/jJAZ1 polypeptides useful as immunogens or
PT antigens to raise or test anti-JAZF1, jJAZ1 or JAZF1/jJAZ1 antibodies.
XX XX
PS Claim 7; Fig 3; 76pp; English.
XX XX
CC The present invention relates to a new JAZF1 (juxtaposed with another
CC zinc finger), jJAZ1 (joined with JAZF1) or JAZF1/jJAZ1 polypeptide. The
CC methods of the invention can be used to identify a compound which
CC controls proliferation of endometrial stroma, by expressing jJAZ1 in the
CC presence of the compound, and determining whether the compound affects
CC expression of jJAZ1, jJAZ1 or JAZF1/jJAZ1 polypeptides are useful
CC as immunogens or antigens to raise or test anti-JAZF1, jJAZ1 or
CC JAZF1/jJAZ1 antibodies. The invention can be used as bait proteins in a
CC two hybrid assay or three hybrid assay to identify other proteins which
CC bind or interact with JAZF1/jJAZ1-binding proteins. JAZF1, jJAZ1 or
CC JAZF1/jJAZ1 molecules are useful for identifying the origin of tumour and
CC as tumour marker protein to verify that a stromal tumour is from
CC endometrium. The antibody is useful for promoting or decreasing fertility
CC or pregnancy, and also for treating endometrial stromal tumours. The
CC present amino acid sequence represents the human JAZF1/jJAZ1 fusion
CC protein of the invention
XX XX
SQ Sequence 776 AA;

Alignment Scores:
Pred. No.: 0 Length: 776
Score: 3725.00 Matches: 726
Percent Similarity: 93.56% Conservative: 0
Best Local Similarity: 93.56% Mismatches: 50
Query Match: 48.44% Indels: 0
DB: 5 Gaps: 0

US-09-874-162A-7 (1-4409) x ABG30843 (1-776)

QY 52 ATGACAGCAGTCGCGCGCGCTCTCTCTCCCAATCCTCCGATCGGGGGCTGGGA 111
Db 1 MetThrGlyIleAlaAlaLaserPhePheSerAsn***** 20

QY 112 CTCACATCTCCACACCTCGCGCGACCTCATCGAGCACATCGAGCAACACATCGATACA 171
Db 21 ***** 40

QY 172 GATCCACGGGTTTAGAAAAACAGAAATACAGCAGCGCAACCTATGTGCCCTGAGTTAC 231
Db 41 AspProArgValLeuGluGlyGlnGluLeuGlnProThrThrValAlaLeuSerIyr 60

QY 232 ATAAATAGATTATGACAGATGCTGCCCGCGAGCAGCAGGAGTCCCTAAAGAAAGAGATT 291
Db 61 IleAsnArgPheMetThrAspAlaAlaArgGluGlnGluSerLeuLysLysLysile 80

QY 292 CAGCGGAGCTCTCGCTGACTCTGTCCAGCTCAGTGTCTCGAGGAGATGTCTCACTCCC 351
Db 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100

QY 352 CCACGCCACAGCAGTGGAGGCTTACTCCCGCGCTGACCCCGACCCATCACCCTCCCTCT 411

101 ProArgHisSerSerGlySerLeuThrProProValThrProProIleThrProSerSer 120
QY 412 TCATTCGCGCAGCAGCAGCTCCGACAGAGCCAAACACAGATCTATAGATTTCTTCGAACCTGG 471
Db 121 SerPheArgSerSerThrProThrGluProThrGlnIleIleArgPheLeuArgThrArg 140
QY 472 AATCTCATAGACCAATATTTTTCGACAGAACTCTTACTTACATGCTCTCATCGAACTCC 531
Db 141 AsnLeuIleAlaProIlePheLeuHisArgThrLeuThrIleMetSerHisArgAsnSer 160
QY 532 AGAACAAACATCAAAAGCGAAACATTTAAAGTTGATGATATCTTATCAAAAGTAGAGAAA 591
Db 161 ArgThrAsnIleLysArgLysThrPheLysValAspAspMetLeuSerLysValGluLys 180
QY 592 ATGAAAGGAGAGCAAGAATCTCATAGCTTGTGAGCTCATTTTCAGCTTACGTTTACTGGT 651
Db 181 MetLysGlyGluGlnGluSerHisSerLeuSerAlaHisLeuGlnLeuThrPheThrGly 200
QY 652 TTCTTCCCAAAATGATAAGCCATCAACAACTCAGAAATGAACAAATTCGTTTACC 711
Db 201 PhePheHisLysAsnAspLysProSerProAsnSerGluAsnGluGlnAsnSerValThr 220
QY 712 CTGGAAGTCTCTGTTGAAAAGTTTGCACAAAAGAAAGAGATGTAAAGTTGTCCTCAATA 771
Db 221 LeuGluValLeuLeuValLysValCysHisLysLysArgLysAspValSerCysProIle 240
QY 772 AGCGAAGTCCACAGGTAAAGCAGGTGCTTTGATTCCTGCTCACTCAATCAACAAACAAA 831
Db 241 ArgGlnValProThrGlyLysGlnValProLeuIleProAspLeuAsnGlnThrLys 260
QY 832 CCGGAAATTTCCGCTCTGCTTCCAGTATGAATTTGAACCTAGTAACAGCCAT 891
Db 261 ProGlyAsnPheProSerLeuAlaValSerAsnGluPheGluProSerAsnSerHis 280
QY 892 ATGTGGAAGTCTTACTGCTGCTTATAGAGTACTGCTCCAGGAGAGAGAGATTAAAT 951
Db 281 MetValLysSerIyrSerLeuLeuPheArgValThrArgProGlyArgArgGluPheAsn 300
QY 952 GGAATGATTAATGAGAAACCAATCAAAATATGTATGTCAATGAGAGCTTCCAGCCAGA 1011
Db 301 GlyMetIleAsnGlyGluThrAsnGluAsnIleAspValAsnGluLeuProAlaArg 320
QY 1012 AGAAAACGAAATCGTGAGATGGGAAAGACATTTGTTGCAAAATACAGATATTTCAT 1071
Db 321 ArgLysArgAsnArgGluAspGlyGlyLysThrPheValAlaGlnMetThrValPheAsp 340
QY 1072 AAAACACGCGCTTACAGCTTTTACATGGGAATATGAAGTAGCATCATCAGCAAAATGAA 1131
Db 341 LysAsnArgArgLeuGlnLeuLeuAspGlyGlyLysThrGluValAlaMet**GluMetGlu 360
QY 1132 GAATGTCCAATGAAGCAAGAAAGACCAATGGGAGACTTCTTTGATGGGAAGAGGTG 1191
Db 361 GluCysProIleSerLysLysArgAlaThrTrpGluThrIleLeuAspGlyLysArgLeu 380
QY 1192 CTTCCATTCGAAACATTTTCTCAGGAGCTAGTTGAGTTTCACTCTTCGTGGACAGA 1251
Db 381 ProProPheGluThrPheSerGlnGlyProThrLeuGlnPheThrLeuArgTrpThrGly 400
QY 1252 GAGACCAATGATAAATCTACGGCTCTTATTCGCAAACTCTTTCGCCACTAGAAATTCAGAG 1311
Db 401 GluThrAsnAspLysSerThrAlaProIleAlaLysProLeuAlaThrArgAsnSerGlu 420
QY 1312 AGTCTCCATCAGAAAAACAGCTGTTTCAGTTTAACTTAACTCAACTATTGCTGTTAAA 1371
Db 421 SerLeuHisGlnGluAsnLysProGlySerValLysProThrGlnThrIleAlaValLys 440
QY 1372 GAATCAATTCACATACAGATCTCAAAACAAAGAAAGAAAGAGTACTCCAAATGAAACCGA 1431
Db 441 GluSerLeuThrThrAspLeuGlnThrArgLysGluLysAspThrProAsnGluAsnArg 460
QY 1432 CAAAAATTAAGAAATATTTTATCAGTTTCTCTATATAACAACAATACAGGCAACAACTGAA 1491
Db 461 GlnLysLeuArgIlePheThrGlnPheLeuThrAsnAsnThrArgGlnGlnThrGlu 480

| | | | | | | | |
|----|------|---|------|----------|---|---|------|
| QY | 856 | GTTCACGTAATGAATTTGAACCTAGTAACGCCATATGGTGAAGTCTTACTCGTTCGTA | 915 | QY | 1936 | CTAAGAGAAAAACCATTAACAAATTAAGAGTCTTTCTGATGTTAAATGAAGAGAGAAA | 1995 |
| DB | 296 | ValSerAsnGluPheGluProSerAsnSerHisMetValLysSerTyrSerLeuLeu | 315 | DB | 656 | LeuArgGluLysThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluLys | 675 |
| QY | 916 | TTTAGAGTACTCGTCACGAGAGAGAGAGTATTAATGAATGATTAATGGAGAAACCAAT | 975 | QY | 1996 | GAAGTGAATGAATCTCGGAATCTCATGTGATGAAGCATGGGTATTATTCGTGACAAATCAA | 2055 |
| DB | 316 | PheArgValThrArgProGlyArgArgGluPheAsnGlyMetIleAsnGlyGluThrAsn | 335 | DB | 676 | GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln | 695 |
| QY | 976 | GAAATATGTCTCAATGAAGAGCTCCAGCCAGAGAAACCAATCGTGAAGATGGG | 1035 | QY | 2056 | ATGAATCATGCTGTATGCTGTTGTAGAAAAATTAAGACAGAAAAATTAATTAAGAGAAT | 2115 |
| DB | 336 | GluAsnIleAspValAsnGluGluLeuProAlaArgArgLysArgAsnArgGluAspGly | 355 | DB | 696 | MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleIleLysLysAsn | 715 |
| QY | 1036 | GAAGAGACATTTGTCACAAATGACAGTATTTGATAAAAAACAGCGCTTACAGCTTTTA | 1095 | QY | 2116 | TTATGTCGAAATTCATGCTTCATCTAGTCAGCATGCATCTTTAAATCTTATTAGCATA | 2175 |
| DB | 356 | GluLysThrPheValAlaGlnMetThrValPheAspLysAsnArgGluLeuLeuLeu | 375 | DB | 716 | LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle | 735 |
| QY | 1096 | GATGGGAATATGAAGTAGCCATGAGGAATGGAGAAATCCAAATAAGCAAGAAAAGA | 1155 | QY | 2176 | ATGTCATATGATAAAGCTGTTACCAAGCTCCGTGAATTCGACCAAAATTAAGAAAAGGG | 2235 |
| DB | 376 | AspGlyGluTyrGluValAlaMetGlnGluMetGluGluCysProIleSerLysLysArg | 395 | DB | 736 | MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly | 755 |
| QY | 1156 | GCAACATGGAGACTATTCTTGATGGAGAGGCTGCTCCATTCGAAAAACATTTCTCAG | 1215 | QY | 2236 | GAATCTGCTTCCCTCCAAACCAAGAAATAAATGAAGAACCAAAATGGGACAGCAAAATGA | 2295 |
| DB | 396 | AlaThrTrpGluThrIleLeuAspGlyLysArgLeuProPheGluThrPheSerGln | 415 | DB | 756 | GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly | 775 |
| QY | 1216 | GGACCTAGCTTGCACTTCCTTCGTTGGACAGAGAGACCAATGATAAATCTACGGCT | 1275 | QY | 2296 | TTTAGTGAATTAATCAAAAGAGAAAGCTTTGGAAAACAGATAGTGTCTCAGGGTTTCA | 2355 |
| DB | 416 | GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAla | 435 | DB | 776 | PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer | 795 |
| QY | 1276 | CCTATTGCGCAACCTCTTGGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAACCAAGCT | 1335 | QY | 2356 | AAACAGAGCAAAAAACAAAACCTC 2379 | |
| DB | 436 | ProIleAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro | 455 | DB | 796 | LysGlnSerLysLysGlnLysLeu 803 | |
| QY | 1336 | GGTTCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAAATCAATGACTACAGATCTCAA | 1395 | RESULT 3 | | | |
| DB | 456 | GlySerValLysProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGln | 475 | ABG30842 | | | |
| QY | 1396 | ACAAGAAAAGAAAGGATCTCCAAATGAACCAACGACCAAAATTAAGAAATTTATCAG | 1455 | ID | ABG30842 | standard; protein; 739 AA. | |
| DB | 476 | ThrArgLysGluLysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyrGln | 495 | AC | ABG30842; | | |
| QY | 1456 | TTTCTCTATAACAATPACAAAGCAACAACTGAAAGCAAGAGATGACCTGCAATGCGCT | 1515 | XX | 21-OCT-2002 | (first entry) | |
| DB | 496 | PheLeuTyrAsnAsnAsnThrArgGlnGlnThrGluAlaArgAspLeuHisCysPro | 515 | XX | Human | joined with jAZF1 (jJAZ1) protein. | |
| QY | 1516 | TGGTGACTCTGAACCTCGCGCAAACTTTATGTTACTCAAGCATCTTAAACTCTGCCAT | 1575 | XX | Human; jAZF1; juxtaposed with another zinc finger; jJAZ1, jAZF1/jJAZ1; | | |
| DB | 516 | TrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuCysHis | 535 | KW | joined with jAZF1; proliferation; endometrial stroma tumour; immunogen; | | |
| QY | 1576 | AGCAGATTATCTCAACTATGTTTATCATCAAAAGTGCTAGATAGATGTTTCTATC | 1635 | KW | antigen; antibody; fertility; pregnancy; gene therapy; vaccine; | | |
| DB | 536 | SerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIle | 555 | KW | chromosome 17. | | |
| QY | 1636 | AATGAGTCTTATGCTCTCTATGCAAGAAATCCTCAGGATATTCATCCCAACCTGGA | 1695 | OS | Homo sapiens. | | |
| DB | 556 | AsnGluCysTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGly | 575 | XX | Key | Location/Qualifiers | |
| QY | 1696 | TTTGCTTTTACTCGCAAGCAGCTTAAGAGACACCTATCACATATTTCTGTGTC | 1755 | XX | 92_93 | | |
| DB | 576 | PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys | 595 | FT | Region | /note="Breakpoint for production of fusion protein | |
| QY | 1756 | AGGCCAAACGAAACAAAGCAGCATGTCTGAATTTCTGAATCTGAAGATGGGAGTA | 1815 | FT | /note="jAZF1/jJAZ1" | | |
| DB | 596 | ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal | 615 | FT | Misc-difference 448..471 | | |
| QY | 1816 | GAAAGCAAGAAACATATAGTAGTGGCCACATGCTGTGATTTTCCATAGTATACCTGC | 1875 | FT | /note="Encoded by CTGCATGCGCTTGGTGTACTCTGAACCTGCCGCAA- | | |
| DB | 616 | GluGlnGlnArgThrTyrSerSerGlyHisAsnArgLeuTyrPheHisSerAspThrCys | 635 | FT | ACTTATAGTATTACTCAAGCATCTAAACTCTGCCAT. C2H2 zinc finger | | |
| QY | 1876 | TTACTCTCCGTCACAAAGAAATGGAAGTAGATAGTGAAGTGAAGTGAAGTGAAGTGAAGT | 1935 | FT | domain" | | |
| DB | 636 | LeuProLeuArgProGlnGlnGluMetGluValAspSerGluAspGluLysAspProGluTyr | 655 | FT | 521..538 | | |
| | | | | FT | /note="Bipartite nuclear localisation signal" | | |
| | | | | XX | WO200193805-A2. | | |
| | | | | XX | 13-DEC-2001. | | |
| | | | | XX | 04-JUN-2001; 2001WO-US017936. | | |
| | | | | XX | 02-JUN-2000; 2000US-0209093P. | | |
| | | | | XX | (BGHM) BRIGHAM & WOMENS HOSPITAL INC. | | |
| | | | | XX | Koontz J, Sklar J; | | |
| | | | | XX | WPI; 2002-575047/61. | | |

Db 632 MetAsnHisAlaCysMetLeuPheValGluAsnTyrglyGlnLysIlelleLysLysAsn 651
QY 2116 TTATGTCGAACACTTCATGCTTCATGTCAGTCATGCTTATCTTATTTAGCAT 2175
Db 652 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 671
QY 2176 ATGTCATAGATAAAGCTGTTACCAAGCTCCGTGAAATGCAGCAAAATAGAAAAGGGG 2235
Db 672 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 691
QY 2236 GAATGCTTCCCTGCAACAGCAAAATACTGAAGCAAAAATGGACAGCAAAATGGA 2295
Db 692 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 711
QY 2296 TTTAGTGAATAAATACTCAAAAGAGAAAGCTTTGGAAACAGATAGTCTCAGGGGTTTCA 2355
Db 712 PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 731
QY 2356 AAACAGACAAACAAACAAACTC 2379
Db 732 LysGlnSerLysLysGlnLysLeu 739
RESULT 4
AAU15978
ID AAU15978 standard; protein; 388 AA.
AC AAU15978;
DT 07-NOV-2001 (first entry)
DE Human novel secreted protein, Seq ID 931.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX Homo sapiens.
XX
FN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217436P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241825P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.

CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Alignment Scores:
 Pred. No.: 2,22e-192 Length: 388
 Score: 2043.00 Matches: 386
 Percent Similarity: 99.48% Conservative: 0
 Best Local Similarity: 99.48% Mismatches: 2
 Query Match: 26.57% Indels: 0
 DB: 4 Gaps: 0

US-09-874-162A-7 (1-4409) x AAU15978 (1-388)

| | | | |
|----|------|---|------|
| QY | 1216 | GGACCTACGTTGCAGTTCCTCACTCTTCGTTGGACGAGAGACCAATGATAATCTACGGCT | 1275 |
| Db | 1 | GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAla | 20 |
| QY | 1276 | CCTATTGCCAAACCTCTTGGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAAAACAAGCCT | 1335 |
| Db | 21 | ProIleAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro | 40 |
| QY | 1336 | GGTTCAGTTAAACCTACTCAAACTATTGCTGCTTAAAGAATCATTTAGACTACAGATCTACAA | 1395 |
| Db | 41 | GlySerValLysProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLysGln | 60 |
| QY | 1396 | ACAAGAAAAAGAAAGGATACTCCAAATGAAAAACCCAGCAAAATTAAGAATATTTATCAG | 1455 |
| Db | 61 | ThrArgLysGluLysAspThrProAsnGluAsnArgGlnLysLeuAlaGliePheTyGln | 80 |
| QY | 1456 | TTTCTCTATACAAATACAGGACCAAACTCAAGCAGAGATGACCTGCATGCGCCT | 1515 |
| Db | 81 | PheLeuTyThrAsnAsnAsnThrArgGlnGlnThrGluAlaArgAspAspLeuHisCysPro | 100 |
| QY | 1516 | TGGTGTACTCTCAACTGCGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCTGCGCAT | 1575 |
| Db | 101 | TrpCysThrLeuAsnCysArgLysLeuTySerLeuLeuLysHisLeuLysLeuCysHis | 120 |
| QY | 1576 | AGCAGATTATCTTCAACTATGTTTATCATCCAAAAGTGTCTAGATAGATGTTTCTATC | 1635 |
| Db | 121 | SerArgPheIlePheAsnTyValTyHisProLysGlyAlaArgIleAspValSerIle | 140 |
| QY | 1636 | AATGAGTGTATGATGGCTCCTATGCAGGAAATCCTCAGAGATATTCATCGCAACCTGGA | 1695 |
| Db | 141 | AsnGluGlyTyArgGlySerTyAlaGlyAsnProGlnAspIleHisArgGlnProGly | 160 |
| QY | 1696 | TTTGTCTTTAGTCGCAACGGACCACTTAAGAGAACACTATTCACACATATTTGTGTGC | 1755 |
| Db | 161 | PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys | 180 |
| QY | 1756 | AGGCCAAAACGAAACAAAGCAGCATGTCGAATTTCTGAATCTCGATAGATGGGGAAGTA | 1815 |
| Db | 181 | ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal | 200 |
| QY | 1816 | GAACAGCAAGAAACATATAGTAGTGGCCACAATCGTCTGTATTTTCCCATAGTACCTGC | 1875 |
| Db | 201 | GluGlnGlnArgThrTySerSerGlyHisAsnArgLeuTyPheHisArgAspThrCys | 220 |
| QY | 1876 | TTACTCTCCGTCACAGAAATGGAAGTAGATAGTCAAGATGAAAAGGATCTCGAATGG | 1935 |
| Db | 221 | LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAspProGluTrp | 240 |
| QY | 1936 | CTAAGAGAAAAAACCATFACACAAATGAAGAGTTTTCTGATGTTAATGAAGAGAGAAA | 1995 |
| Db | 241 | LeuArgGluLysThrIleThrGlnIleGluPheSerAspValAsnGluGlyGluLys | 260 |
| QY | 1996 | GAAGTGTAAACCTCTGGAATCTCCATGTCATGAAGCATGGTTTTATTGTGCAATCAA | 2055 |
| Db | 261 | GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln | 280 |

08-NOV-2000; 2000US-0246525P.
 08-NOV-2000; 2000US-0246526P.
 08-NOV-2000; 2000US-0246527P.
 08-NOV-2000; 2000US-0246528P.
 08-NOV-2000; 2000US-0246529P.
 08-NOV-2000; 2000US-0246530P.
 08-NOV-2000; 2000US-0246531P.
 08-NOV-2000; 2000US-0246532P.
 08-NOV-2000; 2000US-0246533P.
 08-NOV-2000; 2000US-0246534P.
 08-NOV-2000; 2000US-0246535P.
 08-NOV-2000; 2000US-0246536P.
 08-NOV-2000; 2000US-0246537P.
 08-NOV-2000; 2000US-0246538P.
 08-NOV-2000; 2000US-0246539P.
 08-NOV-2000; 2000US-0246540P.
 08-NOV-2000; 2000US-0246541P.
 08-NOV-2000; 2000US-0246542P.
 08-NOV-2000; 2000US-0246543P.
 08-NOV-2000; 2000US-0246544P.
 08-NOV-2000; 2000US-0246545P.
 08-NOV-2000; 2000US-0246546P.
 08-NOV-2000; 2000US-0246547P.
 08-NOV-2000; 2000US-0246548P.
 08-NOV-2000; 2000US-0246549P.
 08-NOV-2000; 2000US-0246550P.
 08-NOV-2000; 2000US-0246551P.
 08-NOV-2000; 2000US-0246552P.
 08-NOV-2000; 2000US-0246553P.
 08-NOV-2000; 2000US-0246554P.
 08-NOV-2000; 2000US-0246555P.
 08-NOV-2000; 2000US-0246556P.
 08-NOV-2000; 2000US-0246557P.
 08-NOV-2000; 2000US-0246558P.
 08-NOV-2000; 2000US-0246559P.
 08-NOV-2000; 2000US-0246560P.
 08-NOV-2000; 2000US-0246561P.
 08-NOV-2000; 2000US-0246562P.
 08-NOV-2000; 2000US-0246563P.
 08-NOV-2000; 2000US-0246564P.
 08-NOV-2000; 2000US-0246565P.
 08-NOV-2000; 2000US-0246566P.
 08-NOV-2000; 2000US-0246567P.
 08-NOV-2000; 2000US-0246568P.
 08-NOV-2000; 2000US-0246569P.
 08-NOV-2000; 2000US-0246570P.
 08-NOV-2000; 2000US-0246571P.
 08-NOV-2000; 2000US-0246572P.
 08-NOV-2000; 2000US-0246573P.
 08-NOV-2000; 2000US-0246574P.
 08-NOV-2000; 2000US-0246575P.
 08-NOV-2000; 2000US-0246576P.
 08-NOV-2000; 2000US-0246577P.
 08-NOV-2000; 2000US-0246578P.
 08-NOV-2000; 2000US-0246579P.
 08-NOV-2000; 2000US-0246580P.
 08-NOV-2000; 2000US-024

QY 2056 ATGAATCATGCTGATGCTGTTGTAGAAAATTATGCACAGAAAATAATTAAAGAAAT 2115
 Db 281 MetAsnHisAlaCysMetLeuPheValGluAsnFyrGlyGlnLysIleIleLysLysAsn 300
 QY 2116 TTATGTCGAATCACTTCATCTATCTAGTCAGCATGTCATGCTTTAATCTTATTAGCATA 2175
 Db 301 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 320
 QY 2176 ATGTCATAGATAAGCTGTTACCAAGCTCCGTGAATATGCAGCAAAAATTAGAAAACGGG 2235
 Db 321 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 340
 QY 2236 GAATCTGCTTCCCTGCAACAGAAAGAAATACTCAAGAAACAAAATGGAGCAGCAAAATGGA 2295
 Db 341 GluSerAlaSerProAlaAsnGluGluIleThrGluGlnAsnGlyThrAlaAsnGly 360
 QY 2296 TTTAGTGAATTAATACTCAAAAGAGAAAGCTTTGGAAAAGATAGTGTCTCAGGGGTTTCA 2355
 Db 361 PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 380
 QY 2356 AAACAGACGCAAAAACAAACTC 2379
 Db 381 LysGlnSerLysLysGlnLysLeu 388

RESULT 5

ABU55047
 ID ABU55047 standard; protein; 388 AA.

| | | | | |
|---|------------------------|----|----|--------------------------------|
| XX AC ABU55047; | PA (ROSE/) ROSEN C A. | XX | PI | Rosen CA, Ruben SM, Barash SC; |
| XX XX | PA (RUBE/) RUBEN S M. | XX | PI | |
| XX DT 18-MAR-2003 (first entry) | PA (BARA/) BARASH S C. | XX | PI | |
| XX DE Human novel polypeptide #134. | | XX | PI | |
| XX KW Human; neural disorder; immune system disorder; renal disorder; | | XX | PI | |
| XX KW muscular disorder; respiratory disease; reproductive disorder; | | XX | PI | |
| XX KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder; | | XX | PI | |
| XX KW hyperproliferative disorder; inflammatory disease; allergic reaction; | | XX | PI | |
| XX KW blood related disorder; cancer; immunosuppressive; antiinflammatory; | | XX | PI | |
| XX KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic; | | XX | PI | |
| XX KW haemostatic; antiarteriosclerotic. | | XX | PI | |
| XX OS Homo sapiens. | | XX | PI | |
| XX FN US2002132753-A1. | | XX | PI | |
| XX PD 19-SEP-2002. | | XX | PI | |
| XX PF 17-JAN-2001; 2001US-00764864. | | XX | PI | |
| XX PR 31-JAN-2000; 2000US-0179065P. | | XX | PI | |
| XX PR 04-FEB-2000; 2000US-0180628P. | | XX | PI | |
| XX PR 28-JUN-2000; 2000US-0214886P. | | XX | PI | |
| XX PR 07-JUL-2000; 2000US-0216647P. | | XX | PI | |
| XX PR 07-JUL-2000; 2000US-0216880P. | | XX | PI | |
| XX PR 11-JUL-2000; 2000US-0217487P. | | XX | PI | |
| XX PR 11-JUL-2000; 2000US-0217496P. | | XX | PI | |
| XX PR 14-JUL-2000; 2000US-0218290P. | | XX | PI | |
| XX PR 26-JUL-2000; 2000US-0220963P. | | XX | PI | |
| XX PR 26-JUL-2000; 2000US-0220964P. | | XX | PI | |
| XX PR 14-AUG-2000; 2000US-0224518P. | | XX | PI | |
| XX PR 14-AUG-2000; 2000US-0224519P. | | XX | PI | |
| XX PR 14-AUG-2000; 2000US-0225267P. | | XX | PI | |
| XX PR 14-AUG-2000; 2000US-0225268P. | | XX | PI | |
| XX PR 14-AUG-2000; 2000US-0225270P. | | XX | PI | |
| XX PR 14-AUG-2000; 2000US-0225447P. | | XX | PI | |
| XX PR 14-AUG-2000; 2000US-0225757P. | | XX | PI | |
| XX PR 22-AUG-2000; 2000US-0225758P. | | XX | PI | |
| XX PR 30-AUG-2000; 2000US-0226868P. | | XX | PI | |
| XX PR 01-SEP-2000; 2000US-0228924P. | | XX | PI | |
| XX PR 01-SEP-2000; 2000US-0229287P. | | XX | PI | |
| XX PR 01-SEP-2000; 2000US-0229343P. | | XX | PI | |
| XX PR 01-SEP-2000; 2000US-0229344P. | | XX | PI | |

QY 2056 ATGAATCATGCTGATGCTGTTGTAGAAAATTATGCACAGAAAATAATTAAAGAAAT 2115
 Db 281 MetAsnHisAlaCysMetLeuPheValGluAsnFyrGlyGlnLysIleIleLysLysAsn 300
 QY 2116 TTATGTCGAATCACTTCATCTATCTAGTCAGCATGTCATGCTTTAATCTTATTAGCATA 2175
 Db 301 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 320
 QY 2176 ATGTCATAGATAAGCTGTTACCAAGCTCCGTGAATATGCAGCAAAAATTAGAAAACGGG 2235
 Db 321 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 340
 QY 2236 GAATCTGCTTCCCTGCAACAGAAAGAAATACTCAAGAAACAAAATGGAGCAGCAAAATGGA 2295
 Db 341 GluSerAlaSerProAlaAsnGluGluIleThrGluGlnAsnGlyThrAlaAsnGly 360
 QY 2296 TTTAGTGAATTAATACTCAAAAGAGAAAGCTTTGGAAAAGATAGTGTCTCAGGGGTTTCA 2355
 Db 361 PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 380
 QY 2356 AAACAGACGCAAAAACAAACTC 2379
 Db 381 LysGlnSerLysLysGlnLysLeu 388

ABU55047
 ID ABU55047 standard; protein; 388 AA.

XX AC ABU55047;

XX XX

XX DT 18-MAR-2003 (first entry)

XX DE Human novel polypeptide #134.

XX KW Human; neural disorder; immune system disorder; renal disorder;
 muscular disorder; respiratory disease; reproductive disorder;
 gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 hyperproliferative disorder; inflammatory disease; allergic reaction;
 blood related disorder; cancer; immunosuppressive; antiinflammatory;
 cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 haemostatic; antiarteriosclerotic.

XX OS Homo sapiens.

XX FN US2002132753-A1.

XX PD 19-SEP-2002.

XX PF 17-JAN-2001; 2001US-00764864.

XX PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 22-AUG-2000; 2000US-0225758P.
 PR 30-AUG-2000; 2000US-0226868P.
 PR 01-SEP-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.

QY 1216 GGACCTAGCTTCAGTTCCTCTTCGTTGGAGGAGACCAATGATAATCTACGCT 1275

Alignment Scores:
 Pred. No.: 2,22e-192 Length: 388
 Score: 2043.00 Matches: 386
 Percent Similarity: 99.48% Conservative: 0
 Best Local Similarity: 99.48% Mismatches: 2
 Query Match: 26.57% Indels: 0
 DB: Gaps: 0

US-09-874-162A-7 (1-4409) x ABU55047 (1-388)

Db 1 GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAla 20
QY 1276 CTTATTGCCAACTCTGCGCACTAGAAATTCAGAGAGTCTCCATCAGGAAACAGAGCCT 1335
Db 21 ProIleAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro 40
QY 1336 GGTTCAGTTAAACCTACTCAAACTATTGCTGTATAAGAAATCATCTACAGATCTACAA 1395
Db 41 GlySerValLysProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGln 60
QY 1396 ACAAGAAAGAAAGAGTACTCCAAATGAAGAACCCACAAATAATTAAGAAATATTATCAG 1455
Db 61 ThrArgLysGluLysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyrGln 80
QY 1456 TTCTCTATAACAACTACAGGCAACAACTGAAGCAAGAGATGACCTGCTGCTGCCT 1515
Db 81 PheLeuTyrAsnAsnThrArgGlnGlnThrGluAlaArgAspAspLeuHisCysPro 100
QY 1516 TGGTGTACTGAACTGCGGCAACTTTATAGTTTACTCAAGCATCTTAAACTCTGCCAT 1575
Db 101 TrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuCysHis 120
QY 1576 AGCAGATTATCTTCAACTATCTTTATCATCCAAAGGTGCTAGGATAGATGTTTCTATC 1635
Db 121 SerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIle 140
QY 1636 AATGAGTGTATGATGGCTCTATCCAGCAAACTCTCAGGATATTTCATCGCAACCTGGA 1695
Db 141 AsnGluGlyTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProgly 160
QY 1696 TTTGCTTTTAGTCGCAACGGACCACTTAAGAGAACACCTATCACACATATTCTTGTGTC 1755
Db 161 PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys 180
QY 1756 AGGCCAAACGAAACAAAGCAAGCATGCTGAATTTCTTGAATCTGAAGATGGGGAAGTA 1815
Db 181 ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal 200
QY 1816 GAACAGCAAAAGCAATATAGTAGTGGCCACAACTCTGTATTTCCATAGTATACCTGC 1875
Db 201 GluGlnGlnArgThrTyrSerSerGlyHisAsnArgLeuTyrPheHisArgAspThrCys 220
QY 1876 TTACTCTCCGTCACAGAATGAAGTATAGTATGATGAGATGAAGAGATCCCTGAATGG 1935
Db 221 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAspProGluTrp 240
QY 1936 CTAAGAGAAACCAACATTACACAAATTCGAAGATTTTCTGATGTTAATGAAGGAGAGAAA 1995
Db 241 LeuArgGluLysThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluLys 260
QY 1996 GAAGTGAATGAACCTCTGCAATCTCATGTCATGAGCATGGTTTATTGCTGACAAATCAA 2055
Db 261 GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 280
QY 2056 ATGAATCATGCTGTATGCTGTTTGTAGAAAATATGACAGAAAATTAATTAAGAGAAAT 2115
Db 281 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleIleLysLysAsn 300
QY 2116 TTATGTCGAACTTCATGCTTCATCTAGTCAGCATGCACTTTAATCTTATTAGCATATA 2175
Db 301 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 320
QY 2176 ATGTCAATAGATAAAGCTGTTTACCAAGCTCCGTGAATTCGACGAAAAATTAAGAAAGGG 2235
Db 321 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 340
QY 2236 GAATCTGCTCCCTGCCAACGAAGAAATAACTGAAGAACAAAAATGGGACAGCAAAATGGA 2295
Db 341 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 360
QY 2296 TTTAGTGAATTAATCAAAAGAGAAAGCTTTGGAACAGATAGTGTCTCAGGGTTTCA 2355

Db 361 PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 380
QY 2356 AAACAGAGCAAAAAACAAACAACTC 2379
Db 381 LysGlnSerLysLysGlnLysLeu 388
RESULT 6
AAU15958
ID AAU15958 standard; protein; 289 AA.
XX AAU15958;
AC AAU15958;
XX 07-NOV-2001 (first entry)
DT Human novel secreted protein, Seq ID 911.
DE
XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX Homo sapiens.
OS
XX WO200155322-A2.
PN
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US0001341.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184564P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215335P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220363P.
PR 26-JUL-2000; 2000US-0220364P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0227009P.
PR 01-SEP-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 06-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230439P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241825P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.

PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250191P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251983P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.

XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-488783/53.
 DR N-PSDB; AAS25945.
 XX
 XX New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX
 PS Claim 11; SEQ ID NO 911; 980pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiodysplasia,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, and to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Alignment Scores:
 Pred. No.: 1.9e-138 Length: 289
 Score: 1497.00 Matches: 283
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.65% Mismatches: 0
 Query Match: 19.47% Indels: 0
 DB: 4 Gaps: 0

US-09-874-162A-7 (1-4409) x AAU15958 (1-289)

PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.

PS Claim 11; SEQ ID NO 911; 402pp; English.

XX The invention relates to human novel polypeptides and their associated
XX polynucleotides. The polypeptides and polynucleotides are useful in gene
XX therapy for treating, inhibiting or preventing neural disorders, immune
XX system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
XX and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
XX nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
XX gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
XX (e.g. congenital heart defects, Epstein's anomaly and hypoplastic left
XX heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
XX renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
XX leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
XX appendicitis), allergic reactions and conditions (e.g. asthma), blood
XX related disorders (e.g. thrombosis, atherosclerosis and myocardial
XX infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and
XX CC ABUS5748 represent human novel polypeptides of the invention

SQ Sequence 289 AA;

Alignment Scores:
Pred. No.: 1.9e-138 Length: 289
Score: 1497.00 Matches: 283
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.65% Mismatches: 0
Query Match: 19.47% Indels: 0
DB: 6 Gaps: 0

US-09-874-162A-7 (1-4409) x ABUS5027 (1-289)

QY 1528 AACTCCGCAACTTTTACTAGTTTACTCAAGCATCTTAACTCTGCCATAGCAGATTTATC 1587
DB 6 AspCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuCysHisSerArgPheIle 25
QY 1588 TTCACACTGTTTATCATCCAAAGGCTCAGGATAGATGTTTCTATCAATGCTGTAT 1647
DB 26 PheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIleAsnGluCysTyr 45
QY 1648 GATGCTCTCTGACGAGAAATCCTCAGGATATTCATCCCAACCTGGATTTGCTTTTGT 1707
DB 46 AspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGlyPheAlaPheSer 65
QY 1708 CGACGACGACGTTAAGAGACACCTTATCACAATATTTCTGTGTCAGGCCCAACGA 1767
DB 66 ArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCysArgProLysArg 85
QY 1768 ACAAGAAGCAAGCATGCTGAATTTCTGAATCTGAAGATGGGAGTAGAACAGCAAGA 1827
DB 86 ThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluValGluGlnArg 105
QY 1828 ACATATAGTAGGCGCAACATCGTGTATTTCCATAGTAGTACTGCTTACCTCTCCGT 1887
DB 106 ThrTyrSerSerGlyHisArgLysLeuTyrPheHisSerAspThrCysLeuProLeuArg 125
QY 1888 CCACAAGAAATGGAAGTAGATAGTGAAGATCAAAAGGATCCTGAATGCTAAGAGAAAA 1947
DB 126 ProGlnGluMetGluValAspSerGluAspGluLysAspProGluTyrLeuArgGluLys 145
QY 1948 ACCATTACAAATTAAGAGTTTCTGATTTAATGAAGAGAGAAAGATGATGAAA 2007
DB 146 ThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluLysGluValMetLys 165
QY 2008 CTCGTGAACTCCATGTCATGACATGGTGGTTTATGCTGACAAATCAATGAATCATGCC 2067
DB 166 LeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGlnMetAsnHisAla 185
QY 2068 TGTATGCTGTTTGTAGAAAAATTTAGCAGAAAAATTAAGAGAAATTTATGTCGAAAC 2127

DB 186 CysMetLeuPheValGluAsnTyrGlyGlnLysIleIleLysLysAsnLeuCysArgAsn 205
QY 2128 TTCATGCTTTTCATCTAGTCAGCATGTCATGATCTTAACTTATTAGCATATGTCATAGAT 2187
DB 206 PheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIleMetSerIleAsp 225
QY 2188 AAAGCTGTTTACCAAGCTCGTGAATGCGCAAAATAGAAAAGGGGAATCTGCTTCC 2247
DB 226 LysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGlyGlnSerAlaSer 245
QY 2248 CTTGCAAAACCAAGAAATTAACCAAGAACAAATGGACAGCAAAATGGAATTTAGTCAAAAT 2307
DB 246 ProAlaAsnGluGluIleThrGluGlnAsnGlyThrAlaAsnGlyPheSerGluIle 265
QY 2308 AACTCAAAAGAGAAAGCTTTGGAAAACAGATAGTCTCAGGGGTTTCAAAACAGAGCAA 2367
DB 266 AsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSerLysGlnSerLys 285
QY 2368 AAACAAAACATC 2379
DB 286 LysGlnLysLeu 289

RESULT 8
AAU16416
ID AAU16416 standard; protein; 292 AA.
XX AC AAU16416;
XX DT 07-NOV-2001 (first entry)
XX DE Human novel secreted protein, Seq ID 1369.
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; optalmallogical; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
OS Homo sapiens.
XX WO200155322-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001341.
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 11-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232969P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232633P.
PR 14-SEP-2000; 2000US-0232644P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-488783/53.
XX N-PSDB; AAS26403.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX used as food additives or preservatives.
XX
XX Claim 11; SEQ ID NO 1369; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
XX encoded secreted proteins. The nucleic acids and proteins are used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX in diagnosing a pathological condition or susceptibility to a
XX pathological condition. Antibodies to the proteins can also be used in
XX alleviating symptoms associated with the disorders and in diagnostic
XX immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
XX (ELISA). Disorders which are diagnosed or treated include autoimmune
XX diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
XX neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
XX arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
XX bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
XX and many other disorders listed in the specification. The polypeptides
XX can also be used to aid wound healing and epithelial cell proliferation,
XX to prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used

CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Alignment Scores:
 Pred. No.: 3,83e-134 Length: 292
 Score: 1453.50 Matches: 280
 Percent Similarity: 96.25% Conservative: 2
 Best Local Similarity: 95.56% Mismatches: 9
 Query Match: 18.90% Indels: 2
 DB: 4 Gaps: 1

US-09-874-162A-7 (1-4409) x AAU16416 (1-292)

| | | | |
|----|------|---|------|
| QY | 748 | AGAAAGGATGAAGTGTCCAAATAGGCAAGTCCACAGGTAAAGACAGGTCCTTTG | 807 |
| Db | 1 | LysLysAspValSerCysProIle***GlnValProThrGlyLys***Gln***ProLeu | 20 |
| QY | 808 | ATTCCTGACCTCAATCAACAAACCCCGAAATTCCTGCTCCCTGCGAGTTCCAGTAAT | 867 |
| Db | 21 | AsnProAspLeuAsnGlnThrLysProGly***PheProSerLeuAlaValSerSerAsn | 40 |
| QY | 868 | GAAATTTGACCTAGTAACAGCCATATGCTGAAGCTTACTCGTTGCTATTAGAGTGACT | 927 |
| Db | 41 | GluPheGluProSer***SerHisMetValLysSerTyrSerLeuLeuPheArgValThr | 60 |
| QY | 928 | COTCCAGGAAGAAGAGAGATTAAATGGAATGATTAATGAGAAACCAATGAAATATTGAT | 987 |
| Db | 61 | ArgProGlyArgGluPheAsnGlyMetIleAsnGlyGluThrAsnGluAsnIleAsp | 80 |
| QY | 988 | GTCATGAAGAGCTCCAGCAGAGAAAGAAAGAAATCGTAGGATGGGAAAGACATTT | 1047 |
| Db | 81 | ValAsnGluGluLeuProAlaArgLysArgLysArgAsnArgGluAspGlyGluLysThrPhe | 100 |
| QY | 1048 | GTTCACAAATACAGATATTGTAATAAACAAGGGCTTACAGCTTTTAGATGGGGAATAT | 1107 |
| Db | 101 | ValAlaGlnMetThrValPheAspLysAsnArgLysGlnLeuLeuAspGlyGlyTyr | 120 |
| QY | 1108 | GAAGTAGCCATCGAGAAATGGAAGATGTCCAATAAGCAAGAAAGACCAATGGGAG | 1167 |
| Db | 121 | ***ValAlaMetGlnGluMetGluGluCysProIleSerLysLysArgAlaThrTrpGlu | 140 |
| QY | 1168 | ACTATCTTGTGGAGAGAGCTGCTCCATTCGAACATTTCTCAGGGACCTACGTTG | 1227 |
| Db | 141 | ThrIleLeuAspGlyLysArgLysProPheGluThrPheSerGlnGlyProThrLeu | 160 |
| QY | 1228 | CAGTTCACTCTTGTGGAGAGAGACCAATGATTAATCTACGGTCTCTATTGCCAAA | 1287 |
| Db | 161 | GlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAlaProIleAlaLys | 180 |
| QY | 1288 | CCTCTTGCCTAGAAATTCAGAGCTCTCCATCAGGAAACACACCTGTTGAGTTAAA | 1347 |
| Db | 181 | ProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysProGlySerValLys | 200 |
| QY | 1348 | CCTACTCAAACTATTGCTGTTAAAGAATCATTTGACTACAGATCTACAAACAAAGAAAGAA | 1407 |
| Db | 201 | ProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGlnThrArgLysGlu | 220 |
| QY | 1408 | AAGGATCTCCAAATGAAACCGCAAAATAGATATTTTATCAGTTCTCTATAAC | 1467 |
| Db | 221 | LysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyrGlnPheLeuTyrAsn | 240 |
| QY | 1468 | AACATACAGCAACAACTGAAGCAAGAGATGACCTGCTGCTGCTGCTGCTGCTGCTG | 1527 |
| Db | 241 | AsnAsnThrArgGlnGlnThrGluAlaArgAspLysLeuHisCysProTrpCysThrLeu | 260 |
| QY | 1528 | AACGCCGCAACTTTATGTTTACTCAAGCATCTTAACTCTGCCATAGCAGATTTATC | 1587 |
| Db | 261 | AsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuCysHisSerArgPheIle | 280 |
| QY | 1588 | TTCRACTATGTTATC---ATCCAAAGGTCGTAGGA | 1621 |

| | | | |
|----------|----------|---|-----|
| Db | 281 | PheAsn***Val-LeuSerSerGlnLysValLeuGly | 292 |
| RESULT | 9 | | |
| ABU55485 | | | |
| ID | ABU55485 | standard; protein; 292 AA. | |
| XX | AC | ABU55485; | |
| XX | DT | 18-MAR-2003 (first entry) | |
| XX | DE | Human novel polypeptide #572. | |
| XX | KW | Human; neural disorder; immune system disorder; renal disorder; | |
| KW | | muscular disorder; respiratory disease; reproductive disorder; | |
| KW | | gastrointestinal disorder; pulmonary disorder; cardiovascular disorder; | |
| KW | | hyperproliferative disorder; inflammatory disease; allergic reaction; | |
| KW | | blood related disorder; cancer; immunosuppressive; antiinflammatory; | |
| KW | | cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic; | |
| XX | | haemostatic; antiarteriosclerotic. | |
| OS | | Homo sapiens. | |
| PN | | US2002132753-A1. | |
| PD | | 19-SEP-2002. | |
| XX | | | |
| PF | | 17-JAN-2001; 2001US-00764864. | |
| XX | | | |
| PR | | 31-JAN-2000; 2000US-0179065P. | |
| PR | | 04-FEB-2000; 2000US-0180628P. | |
| PR | | 28-JUN-2000; 2000US-0214886P. | |
| PR | | 07-JUL-2000; 2000US-0216647P. | |
| PR | | 07-JUL-2000; 2000US-0216880P. | |
| PR | | 11-JUL-2000; 2000US-0217487P. | |
| PR | | 11-JUL-2000; 2000US-0217496P. | |
| PR | | 14-JUL-2000; 2000US-0218290P. | |
| PR | | 26-JUL-2000; 2000US-0220963P. | |
| PR | | 26-JUL-2000; 2000US-0220964P. | |
| PR | | 14-AUG-2000; 2000US-0224518P. | |
| PR | | 14-AUG-2000; 2000US-0224519P. | |
| PR | | 14-AUG-2000; 2000US-0225267P. | |
| PR | | 14-AUG-2000; 2000US-0225268P. | |
| PR | | 14-AUG-2000; 2000US-0225270P. | |
| PR | | 14-AUG-2000; 2000US-0225447P. | |
| PR | | 14-AUG-2000; 2000US-0225757P. | |
| PR | | 14-AUG-2000; 2000US-0225758P. | |
| PR | | 22-AUG-2000; 2000US-0226868P. | |
| PR | | 30-AUG-2000; 2000US-0228924P. | |
| PR | | 01-SEP-2000; 2000US-0229287P. | |
| PR | | 01-SEP-2000; 2000US-0229343P. | |
| PR | | 01-SEP-2000; 2000US-0229344P. | |
| PR | | 01-SEP-2000; 2000US-0229345P. | |
| PR | | 05-SEP-2000; 2000US-0229509P. | |
| PR | | 05-SEP-2000; 2000US-0229513P. | |
| PR | | 09-SEP-2000; 2000US-0231413P. | |
| PR | | 21-SEP-2000; 2000US-0234223P. | |
| PR | | 21-SEP-2000; 2000US-0234274P. | |
| PR | | 25-SEP-2000; 2000US-0234997P. | |
| PR | | 27-SEP-2000; 2000US-0235834P. | |
| PR | | 29-SEP-2000; 2000US-0236327P. | |
| PR | | 29-SEP-2000; 2000US-0236367P. | |
| PR | | 29-SEP-2000; 2000US-0236368P. | |
| PR | | 29-SEP-2000; 2000US-0236369P. | |
| PR | | 29-SEP-2000; 2000US-0236370P. | |
| PR | | 02-OCT-2000; 2000US-0236802P. | |
| PR | | 02-OCT-2000; 2000US-0237037P. | |
| PR | | 02-OCT-2000; 2000US-0237038P. | |
| PR | | 02-OCT-2000; 2000US-0237039P. | |
| PR | | 02-OCT-2000; 2000US-0237040P. | |
| PR | | 13-OCT-2000; 2000US-0239935P. | |
| PR | | 20-OCT-2000; 2000US-0240960P. | |
| PR | | 20-OCT-2000; 2000US-0241785P. | |

PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 XX WPI; 2003-147444/14.
 DR N-PSDB; ABX73744.
 XX
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 PT
 XX
 PS Claim 11; SEQ ID NO 1369; 402pp; English.
 XX
 CC The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC (e.g. congenital heart defects, Eisen's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
 CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and
 CC ABUS5748 represent human novel polypeptides of the invention
 XX
 SQ Sequence 292 AA;
 Alignment Scores:
 Pred. No.: 3.83e-134 Length: 292
 Score: 1453.50 Matches: 280
 Percent Similarity: 96.25% Conservative: 2
 Best Local Similarity: 95.56% Mismatches: 9
 Query Match: 18.90% Indels: 2
 DB: 6 Gaps: 1
 US-09-874-162A-7 (1-4409) x ABUS5485 (1-292)
 QY 748 AGAAGGATGAAGTGTCCATAGGCAAGTCCACAGGTAAAGCAGGTGCCTTTG 807
 Db 1 LysLysAspValSerCysProile***GlnValProThrGlyLys***Gln***ProLeu 20
 QY 808 ATTCTGACCTCAATCAACAAACAAACCCGGAAATTTCCCGTCCCTTCCAGTATCAGTAAT 867
 Db 21 AsnProAspLeuAsnGlnThrLysProgly***PheProSerLeuAlaValSerSerAsn 40
 QY 868 GAATTTGAACCTAGTAACAGGCATATGTGTAAGTCTTACTCGTTGCTATTTAGAGTGAAT 927
 Db 41 GluPheGluProSer***SerHisMetValLysSerTySerLeuLeuPheArgValThr 60
 QY 928 CGTCCAGGAAGAAGAGCTTTAATGGAATATTATGAGAAACCAATGAATATTTGAT 987
 Db 61 ArgProGlyArgArgGluPheAsnGlyMetIleAsnGlyGluThrAsnGluAsnIleAsp 80
 QY 988 GTCAATGAGAGCTTCCAGCCAGAGAAACGAAATCGTGAGGATGGGAAAGACATTT 1047
 Db 81 ValAsnGluGluLeuProAlaArgArgLysArgAsnArgGluAspGlyGluLysThrPhe 100
 QY 1048 GTTGACAAATGACAGTATTTGATAAAACAGGCGCTTACAGCTTTAGATGGGAATAT 1107

Db 1101 ValAlaGlnMetThrValPheAspLysAsnArgArgLeuGlnLeuLeuAspGlyGluTyr 120
 QY 1108 GAAGTACCCATGACAGGAAATGGAAGATGTCCATATAGCAAGAAAGACAGCAATGGAG 1167
 Db 121 ***ValAlaMetGlnGluMetGluGluCysProileSerLysLysArgAlaThrTrpGlu 140
 QY 1168 ACTATTCTTGATGGGAAGAGCGCTCCCTCCATTCGAAACATTTCTCAGGGACCTACGTTG 1227
 Db 141 ThrileLeuaspGlyLysArgLeuProPheGluThrPheSerGlnGlyProThrLeu 160
 QY 1228 CAGTTCACTCTTCTGTCGACAGGAGACCAATGATAAATCTACGGCTCCTATTGCCAAA 1287
 Db 161 GlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAlaProIleAlaLys 180
 QY 1288 CCTCTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAGAAACAAAGCCTGCTCAGTTAA 1347
 Db 181 ProLeuAlaThrArgAsnSerGluSerLeuHiGlnGluAsnLysProGlySerVallys 200
 QY 1348 CCTACTCAAACTATTGCTGTTAAAGAAATCAATTGACTACAGATCTACAAAACAGAAAGAA 1407
 Db 201 ProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGlnThrArgLysGlu 220
 QY 1408 AAGGTACTCCAAATGAAACCCGACAAATTAAGATATTTTATCAGTTTCTCTATAAC 1467
 Db 221 LysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyThrGlnPheLeuTyAsn 240
 QY 1468 AACAAATACAAAGGCAACAAACTGAAGCAAGAGATGACCTGCAATTCCTGCTACTCTG 1527
 Db 241 AsnAsnThrArgGlnGlnThrGluAaArgAspAspLeuHisCysProTyThrLeu 260
 QY 1528 AACTGCGCGAACTTTATAGTTTACTCAAGCATCTTAACTGCCATAGCAGATTTATC 1587
 Db 261 AsnCysArgLysLeuTySerLeuLeuLysHisLeuLysLeuLysHisSerArgPheIle 280
 QY 1588 TTCAACTATCTTTATC---ATCCAAAGGTGCTAGGA 1621
 Db 281 PheAsn***Val-LeuSerSerGlnLysValLeuGly 292
 RESULT 10
 ABB71635
 ID ABB71635 standard; protein; 955 AA.
 AC ABB71635;
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 41697.
 XX Drosophila melanogaster polypeptide; cell signalling; insecticide;
 XX pharmaceutical.
 OS Drosophila melanogaster.
 XX WO200171042-A2.
 PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US009231.
 PF 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL15738.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX

PS Disclosure, SEQ ID NO 41697; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB16051), expressed DNA sequences (AB16176-AB16175) and the encoded proteins (AB16173-AB16175). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 955 AA;

Alignment Scores:
Pred. No.: 1,23e-101 Length: 955
Score: 1127.50 Matches: 273
Percent Similarity: 49.46% Conservative: 137
Best Local Similarity: 32.93% Mismatches: 240
Query Match: 14.66% Indels: 179
DB: 4 Gaps: 26

US-09-874-162A-7 (1-4409) x ABB71635 (1-955)

| | | | |
|----|-----|--|-----|
| QY | 244 | ATGACAGATCTGCCCGCGAGACAGAGTCCCTAAAGAGAGATTACGCGGAGCTC | 303 |
| DB | 1 | MetAlaProAlaLysLysArgGluLysAspSer | 11 |
| QY | 304 | TCGCTGACTCTGCTCAGCTCAGTCTCTCGAGGGAAATGTG | 351 |
| DB | 12 | -----AsnProAspGlySerAlaAlaAsnGlyIleLeuGlyLeuThrHisGlyAlaPro | 29 |
| QY | 352 | CCAGCCACACAGTGGAGGCTTACTCCCGCGTACCCGCCACCCACCCCTCCTCT | 411 |
| DB | 30 | AspAlaSerAsnAlaGlySerThrValProThrAlaGluGlyGlnValLysLeuAsn | 49 |
| QY | 412 | TCATTCGCGACGACCTCCG | 453 |
| DB | 50 | GlyHisGlnGlnGlnGlnLeuPheLeuGlnAlaPheGluLysProThrGlnIleTyr | 69 |
| QY | 454 | AGATTCTTCGAATCTCGAATCTCATAGCACCATAATTTTGCACAACTCTTACTAC | 513 |
| DB | 70 | ArgTyrLeuArgAsnArgHisGluThrAsnProIlePheLeuAsnArgThrLeuSerTyr | 89 |
| QY | 514 | ATGTCTCATCAAACTCCAGACAAACATCAAAAGGAAACATTAAAGTTGATGATG | 573 |
| DB | 90 | MetLysGluArgMetSerArgAsnAsnLysLysArgIleSerPheGlnValAsnSerMet | 109 |
| QY | 574 | TTATCAAAAGTAGAAATAAGAGGAGACAGAAATCTCATGCTTGTGCTGCT | 630 |
| DB | 110 | LeuGluSerIle-----ThrGlnLysSerGluAlaValSerGlnAsnTyr | 124 |
| QY | 631 | TTGACAGTCTACGTTTACTGGTTCTTCCACAAAATGATAGCCATCACCAACTCAGAA | 690 |
| DB | 125 | LeuHisValIleTyrAspSerLeuHisGluLysLeuProAlaArgLeuAspAsnGluSer | 144 |
| QY | 691 | AATGAACAA-----AATTCTGTACCTCGAA | 717 |
| DB | 145 | GlyGluAspLeuLeuGlnGlnLeuLeuLysGluAlaGlyGluSerValSerValGlu | 164 |
| QY | 718 | GTCCTGCTTGAAGTTGGCCACAAAAGAGAGATGTAAGTTGTCCTCAATAAGCAA | 777 |
| DB | 165 | ThrThrLeuTyrLysIleThrArgSerLysArgLysAspSerThrLeuAspPheGlnGlu | 184 |
| QY | 778 | GTTCCACAGGTAAACAGGTCCTTGTGATTCCTGACCTCAATCAACAAACCCCGGA | 837 |
| DB | 185 | LeuLeuSerLysCysSerGlnIleValTyrAsnPro-----LysAspArgValGly | 201 |
| QY | 838 | AATTTCCCGTCCCTGGAGTTTCCAGTAATGAATTTGAACT-----AGTAACAGCCATATG | 894 |
| DB | 202 | GluHisAlaThrIleSerIleProLeuGlnThrMetArgProMetGlyGluGlnHisThr | 221 |

| | | | |
|----|------|--|------|
| QY | 895 | GTGAAGTCTTACTCGTTCCTATTAGAGTACTCGTCCAGGAGAGAGAGAGTAAATGGA | 954 |
| DB | 222 | Leu-----TyrLysLeuLeuPheArgIleLys | 230 |
| QY | 955 | ATGATTAAATGAGAAACCAATGAAATAATTGATGTCAATGAGAGCTTCCAGCCAGAGA | 1014 |
| DB | 231 | ValLeuSerProSerThrCysAsn-----AspGluAsnAlaGluThrProPro-----Asn | 247 |
| QY | 1015 | AAACCAAAATCGAGGATGGGAAAGACATTGTCACAAATGACAGATATTGATAAA | 1074 |
| DB | 248 | LysArgSerArgProAsn-----GluLysMetPheGlySerGluLeuIleLeuTyrGluLys | 266 |
| QY | 1075 | AAACAGCGCTACAGCTTTTAGATGGGAATATGAAGTAGCCATGACGAAATGAGAA | 1134 |
| DB | 267 | SerSerGlyPhe-----IleThrGluGlyGluTyrGluAlaMetLeuGlnProLeuAsnSer | 285 |
| QY | 1135 | TGTCCAATA-----AGCAAGAAAGACACATCGGAGACTATTCTTGATGGGAAG | 1185 |
| DB | 286 | ThrSerIleLysSerPheSerProLysLysCysThrTrpGluThrMetProAsp-----Ser | 304 |
| QY | 1186 | AGGCTGCGCT-----CCATTGCAAAATTTCTCAGGAGACCTACGTTGCAGTTCAC | 1236 |
| DB | 305 | TyrIleProLeuSerLeuThrTyrAspValTyrGlnGlnSerProMetLeuLysPheHis | 324 |
| QY | 1237 | CTTCGTTGGACAGGAGACCAATGATAATCTACGGCTCCTATTGCGCAACCTCTTGCC | 1296 |
| DB | 325 | LeuThrLeu-----SerAsnGluGlnLeuProGluMetIleSerAlaProGluLeu | 341 |
| QY | 1297 | ACTAGA-----AATTCAGAGAGCTC | 1317 |
| DB | 342 | GlnArgTyrValGlnHisLeuAspAlaValAlaGluMetAsnTyrAsnAsnAsnAsnTyr | 361 |
| QY | 1318 | CATCAGGAAAAAC-----AAGCCTGGTTCAGTTAAACCTTACTCAACTATT | 1362 |
| DB | 362 | AsnAsnAsnAsnAsnCysSerGlyLeuLysAsnGlySerGlyGlyGlyAsnSerThrVal | 381 |
| QY | 1363 | GCTGTTAAAGAAATCAATTGACTACAGATCTACAAACAGAAAGAAAGATCTCCAAT | 1422 |
| DB | 382 | Cys-----LysThrPro----- | 386 |
| QY | 1423 | GAACCCGACAAAAATTAAGAAATATTTTATCAGTTTCTTATAACAACAATACAGGCAA | 1482 |
| DB | 387 | -----GluHisIleGlnIleValTyrAsnPheMetTyrSerAsnAsnThrArgGln | 403 |
| QY | 1483 | CAAACTGAAGCAAGAGATGACCTGCATTGCGCTTGGTGTACTCTGAACTCGCGAACTT | 1542 |
| DB | 404 | GlnThrGluTyrThrGlnGluLeuAsnCysProTyrCysGlyLeuAspCysLeuArgLeu | 423 |
| QY | 1543 | TATAGTTTACTCAAGCATCTTAACTCTGCCATAGCAGATTTATCTTCAACTATGTTTAT | 1602 |
| DB | 424 | TyrAlaLeuLeuLysHisLeuLysLeuCysHisAlaArgPheAsnPheThrTyrGlnPro | 443 |
| QY | 1603 | CATCCAAAAGTCTAGGATAGATGTTTCTATCAATAGTGTATATGATGGCTCTTATGCA | 1662 |
| DB | 444 | AlaGlySerGlyAlaArgIleAspValThrIleAsnAspAlaTyrAspGlySerTyrAla | 463 |
| QY | 1663 | GAATATCTCAGATATTCAATCGCCAACTGGATTGCTTTTAGTCGCAAC-----GGACCA | 1719 |
| DB | 464 | GlySerProTyrAspLeuAlaGlyProSerGlySerSerPheAlaArgThrCysGlyPro | 483 |
| QY | 1720 | GTTAAGAGAACCACTTATCACATATTCTTGTGTGTCAGGCCCAACCAACAAAGCAAGC | 1779 |
| DB | 484 | ValArgArgThrSerValThrSerLeuMetValCysArgProArgArgGlnLysThrCys | 503 |
| QY | 1780 | ATGCTCTGAATTTCTGAATCTGAAGTGGGAGTAGACAGCAAGCAACATATAGTAGT | 1839 |
| DB | 504 | LeuAspGluPheLeuGluLeuAspGluLeuSerAsnGlnArgSerTyrIleThr | 523 |
| QY | 1840 | GCCCAATCTGCTGCTGTTTCTTCCATAGTATGATCTCTTACCTCTCCGTCACAAAGATG | 1899 |
| DB | 524 | GlyHisAsnArgLeuTyrHisHisThrGluThrCysLeuProValHisProLysGluLeu | 543 |
| QY | 1900 | GAAGTAGATAGTGAAGATGAAAGAGATCTGATGGCTTAAGAGAAAAACCATACACAA | 1959 |

| | | | |
|-----------|----------|---|------|
| Db | 544 | AspIleAepSerGluGlyGluSerAspProLeuTrpLeuArgGlnLysThrIleGlnMet | 563 |
| Qy | 1960 | ATTGAAGAGTTTCGTGATGTTAAATGAAGAGAGAAAGATGATCAAACTCTGGAATCTC | 2019 |
| Db | 564 | IleAspGluPheSerAspValAsnGluGlyGluLysGlnLeuMetLysLeuTrpAsnLeu | 583 |
| Qy | 2020 | CATGTCATGAAGCATGGGTTTATGCTGACAAATCAAAATGAATCATGCTGTATGCTCTTT | 2079 |
| Db | 584 | HisValMetArgHisGlyPheValGlyAspCysGlnLeuProIleAlaCysGluMetPhe | 603 |
| Qy | 2080 | GTAGAAAATTATGGACAGAAAATAATTAAGAAGAAATTATGTGAAAACCTTCATGCTTCAT | 2139 |
| Db | 604 | LeuAspAlaLysGlyThrGluIleValArgLysAsnLeuTyrArgAsnPheIleLeuHis | 623 |
| Qy | 2140 | CTAGTCACATGCATGACTTTAACTTTATTAGCAATAATGTCATAGATAAGCTGTACC | 2199 |
| Db | 624 | MetCysSerLeuPheAspTyrGlyLeuIleAlaGlnThrValTyrLysThrValGln | 643 |
| Qy | 2200 | AAGCTCCGTGAAATGCACAAAAATTAGAAAAGGGGAATCT----- | 2241 |
| Db | 644 | LysLeuGlnGlyLeuLeuSerLysTyrAlaAlaGlyGlnGluLeuMetGlnArgGlnArg | 663 |
| Qy | 2241 | ----- | 2241 |
| Db | 664 | GluGluGlnLeuLysTyrTrpLeuAspValGlyMetHisLysLysGlnGluAspProLys | 683 |
| Qy | 2242 | -----GCTCCCTCGAAACAGAAATAACTGAA----- | 2271 |
| Db | 684 | ThrLeuLysSerProGlnLysProAlaProProAlaAspGlnAlaSerThrSerSerAla | 703 |
| Qy | 2271 | ----- | 2271 |
| Db | 704 | SerThrSerGlySerGlySerGlySerSerMetGlnProProLysArgMetProAla | 723 |
| Qy | 2272 | -----GAACAAAATGGACAGCAAT | 2292 |
| Db | 724 | HisLeuLysArgGlySerAlaAlaSerSerProGlyValGlnSerLysGlyThrGluAsn | 743 |
| Qy | 2293 | GCATTTAGTGAATTAACCTCAAAAGAGAAAGCTTTGGAACACATAGTCTCAGGGGTT | 2352 |
| Db | 744 | GlyThrAsnGlySerAsnSer-----SerSerSerAsnSerLysAsnVal | 758 |
| Qy | 2353 | TCAAAACAGACAAAACAAAACAACTC | 2379 |
| Db | 759 | AlaLysSerAlaAspGlnProLeu | 767 |
| RESULT 11 | | | |
| AAU16402 | | | |
| ID | AAU16402 | standard; protein; 175 AA. | |
| XX | AC | AAU16402; | |
| XX | DT | 07-NOV-2001 (first entry) | |
| XX | DE | Human novel secreted protein, Seq ID 1355. | |
| XX | KW | Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; | |
| XX | KW | cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective; | |
| XX | KW | antibacterial; virucide; fungicide; ophthalmological; vulnery; | |
| XX | KW | secreted protein; rheumatoid arthritis; hyperproliferative disorder; | |
| XX | KW | cardiovascular disorder; cardiac arrest; cerebrovascular disorder; | |
| XX | KW | cerebral ischaemia; angiogenesis; nervous system disorder; | |
| XX | KW | Alzheimer's disease; infection; ocular disorder; corneal infection; | |
| XX | KW | wound healing; epithelial cell proliferation; skin ageing; food additive; | |
| XX | OS | preservative; antiproliferative. | |
| XX | OS | Homo sapiens. | |
| XX | FN | WO200155322-A2. | |
| XX | PD | 02-AUG-2001. | |
| XX | XX | | |

| | | | | |
|----|--------------------------------|-------------------|----|---|
| PR | 29-SEP-2000; | 2000US-0236370P. | PT | New nucleic acid molecules encoding 461 human secreted proteins for |
| PR | 02-OCT-2000; | 2000US-0236802P. | PT | diagnosing, preventing, treating or ameliorating medical conditions and |
| PR | 02-OCT-2000; | 2000US-0237037P. | PT | used as food additives or preservatives. |
| PR | 02-OCT-2000; | 2000US-0237038P. | XX | Claim 11; SEQ ID NO 1355; 980pp; English. |
| PR | 02-OCT-2000; | 2000US-0237039P. | XX | The invention relates to isolated nucleic acid molecules and their |
| PR | 02-OCT-2000; | 2000US-0237040P. | XX | encoded secreted proteins. The nucleic acids and proteins are used to |
| PR | 13-OCT-2000; | 2000US-0239935P. | CC | prevent, treat or ameliorate a medical condition in e.g. humans, mice, |
| PR | 13-OCT-2000; | 2000US-0239937P. | CC | rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used |
| PR | 20-OCT-2000; | 2000US-0240960P. | CC | in diagnosing a pathological condition or susceptibility to a |
| PR | 20-OCT-2000; | 2000US-0241221P. | CC | pathological condition. Antibodies to the proteins can also be used in |
| PR | 20-OCT-2000; | 2000US-0241785P. | CC | alleviating symptoms associated with the disorders and in diagnostic |
| PR | 20-OCT-2000; | 2000US-0241786P. | CC | immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays |
| PR | 20-OCT-2000; | 2000US-0241787P. | CC | (ELISA). Disorders which are diagnosed or treated include autoimmune |
| PR | 20-OCT-2000; | 2000US-0241809P. | CC | diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. |
| PR | 20-OCT-2000; | 2000US-0241826P. | CC | neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac |
| PR | 01-NOV-2000; | 2000US-0244617P. | CC | arrest, cerebrovascular disorders e.g. cerebral ischaemia, angogenesis, |
| PR | 08-NOV-2000; | 2000US-0246474P. | CC | nervous system disorders e.g. Alzheimer's disease, infections caused by |
| PR | 08-NOV-2000; | 2000US-0246475P. | CC | bacteria, viruses and fungi and ocular disorders e.g. corneal infection, |
| PR | 08-NOV-2000; | 2000US-0246476P. | CC | and many other disorders listed in the specification. The polypeptides |
| PR | 08-NOV-2000; | 2000US-0246477P. | CC | can also be used to aid wound healing and epithelial cell proliferation, |
| PR | 08-NOV-2000; | 2000US-0246478P. | CC | to prevent skin aging due to sunburn, to maintain organs before |
| PR | 08-NOV-2000; | 2000US-0246532P. | CC | transplantation, for supporting cell culture of primary tissues, to |
| PR | 08-NOV-2000; | 2000US-0246524P. | CC | regenerate tissues and in chemotaxis. The polypeptides can also be used |
| PR | 08-NOV-2000; | 2000US-0246525P. | CC | as a food additive or preservative to increase or decrease storage |
| PR | 08-NOV-2000; | 2000US-0246526P. | CC | capabilities, fat content, lipid, protein, carbohydrate, vitamins, |
| PR | 08-NOV-2000; | 2000US-0246527P. | CC | minerals, cofactors and other nutritional components. The present |
| PR | 08-NOV-2000; | 2000US-0246528P. | CC | sequence represents a novel secreted protein of the invention. Note: The |
| PR | 08-NOV-2000; | 2000US-0246532P. | CC | sequence data for this patent did not form part of the printed |
| PR | 08-NOV-2000; | 2000US-0246603P. | CC | |
| PR | 08-NOV-2000; | 2000US-0246610P. | CC | |
| PR | 08-NOV-2000; | 2000US-0246611P. | CC | |
| PR | 08-NOV-2000; | 2000US-0246613P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249207P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249208P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249209P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249210P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249211P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249212P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249213P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249214P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249215P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249216P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249217P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249218P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249244P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249245P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249264P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249265P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249297P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249299P. | CC | |
| PR | 17-NOV-2000; | 2000US-0249300P. | CC | |
| PR | 01-DEC-2000; | 2000US-0250160P. | CC | |
| PR | 01-DEC-2000; | 2000US-0250391P. | CC | |
| PR | 05-DEC-2000; | 2000US-0251030P. | CC | |
| PR | 05-DEC-2000; | 2000US-0251988P. | CC | |
| PR | 05-DEC-2000; | 2000US-0256719P. | CC | |
| PR | 06-DEC-2000; | 2000US-0251479P. | CC | |
| PR | 08-DEC-2000; | 2000US-0251856P. | CC | |
| PR | 08-DEC-2000; | 2000US-0251868P. | CC | |
| PR | 08-DEC-2000; | 2000US-0251869P. | CC | |
| PR | 08-DEC-2000; | 2000US-0251989P. | CC | |
| PR | 11-DEC-2000; | 2000US-0251990P. | CC | |
| PR | 05-JAN-2001; | 2001US-02554097P. | CC | |
| XX | (HUMA-) HUMAN GENOME SCI INC. | | XX | |
| PI | Rosen CA, Bazash SC, Ruben SM; | | PI | |
| XX | WPI: 2001-488783/53. | | XX | |
| DR | N-PSDB; AAS26389. | | DR | |
| XX | | | XX | |

| | | | |
|----|------|--|------|
| QY | 1447 | TTTTATCAGTTTCTCTATAACAACAACTACAAGCGCAACAACTGAGCAAGAGATGACCTG | 1506 |
| Db | 1 | PheTyrGln**LeuTyrAsnAsnAsnThrArgGlnGlnThrGluAlaArgAspLeu | 20 |
| QY | 1507 | CATTGCCCTTGTTGTTACTCTGAACTGCGCAACTTTATAGTTTACTCAAGCTCTTAA | 1566 |
| Db | 21 | HisCysProTrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLys | 40 |
| QY | 1567 | CTCTGCCATAGCAGATTATCTTCAACTATGTTTATCATCCAAAGAGGTGCTAGGATAGT | 1626 |
| Db | 41 | LeuCysHisSerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleasp | 60 |
| QY | 1627 | GTTTCTATCAATGAGTGTATGATGCTTCCTATGAGGAATCTTCAGATATTCATGCG | 1686 |
| Db | 61 | ValSerIleAsnGluCysTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArg | 80 |
| QY | 1687 | CAACCTGGATTGCTTTTGTAGTCGCAACGAGCCAGTAAAGAGAACACCTTATCACATATT | 1746 |
| Db | 81 | GlnProGlyPheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIle | 100 |
| QY | 1747 | CTTGTGTGAGCCCAAAACGAAACAAAGCAACATGTCATGTTTCTTGAATCTGAAGAT | 1806 |
| Db | 101 | LeuValCysArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluasp | 120 |
| QY | 1807 | GGGAGTAGACACGACCAAGACATATAGTAGTGGCCCACTCGTCTGTATTTCATAGT | 1866 |
| Db | 121 | GlyGluValGluGlnArgThrThrSerSerGlyHisAsnArg**TyrPheHisSer | 140 |
| QY | 1867 | GATACCTGCTTACCTCTCCGTCACCAAGAAATGGAAGTAGATAGTGAAGATGAAAGAT | 1926 |
| Db | 141 | AspThrCysLeuProLeuArgProGlnGluMetGluValAspSerGluaspGluLysasp | 160 |

| | | | | |
|------------------------|----------|---------------|-----|--|
| Alignment Scores: | | | | |
| Pred. No.: | 1.79e-83 | Length: | 175 | |
| Score: | 940.00 | Matches: | 172 | |
| Percent Similarity: | 97.73% | Conservative: | 0 | |
| Best Local Similarity: | 97.73% | Mismatches: | 3 | |
| Query Match: | 12.22% | Indels: | 1 | |
| DB: | 4 | Gaps: | 0 | |

US-09-874-162A-7 (1-4409) x AAU16402 (1-175)

```
QY 1927 CCTGAATGGCTAAGAGAAARACCATTCACAAATTGAGAGTTT 1972
Db ||||| 161 ProGluTrpLeuArg-***LysProLeuHisLysLeuLysSerPhe 175
RESULT 12
ABU55471
ID ABU55471 standard; protein; 175 AA.
XX AC
XX ABU55471;
XX DT 18-MAR-2003 (first entry)
XX DE Human novel polypeptide #558.
XX KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX OS Homo sapiens.
XX PN US2002132753-A1.
XX PD 19-SEP-2002.
XX PF 17-JAN-2001; 2001US-00764864.
XX PR 31-JAN-2000; 2000US-0179063P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226688P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX PI Rosen CA, Ruben SM, Barash SC;
XX WPI; 2003-147444/14.
DR N-PSDB; ABX73730.
XX New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX PS Claim 11; SEQ ID NO 1355; 402pp; English.
XX CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular lupus erythematosus, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, buritis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC ABU55748 represent human novel polypeptides of the invention
XX SQ Sequence 175 AA;
Alignment Scores:
Pred. No.: 1.79e-83 Length: 175
Score: 940.00 Matches: 172
Percent Similarity: 97.73% Conservative: 0
Best Local Similarity: 97.73% Mismatches: 3
Query Match: 12.22% Indels: 1
DB: 6 Gaps: 0
US-09-874-162A-7 (1-4409) x ABU55471 (1-175)
QY 1447 TTTTATCAGTTTCTCTATAACACATACAAAGCCACAACTGAAGCAGAGATGACCTG 1506
||| 1 PheTyrGln***LeuTyrAsnAsnAsnThrArgGlnGlnThrGluAlaArgAspLeu 20
QY 1507 CATGGCCCTGGTGTTACTCTGAACCTCCGCAAACTTTATAGTTTCTCAAGCATCTTAAA 1566
||| 21 HisCysProTrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLys 40
QY 1567 CTCTGCCATAGCAGATTATCTCAACTATGTTTATCATCCAAAGGTCGTAGGATAGAT 1626
||| 41 LeuCysHisSerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAsp 60
QY 1627 GTTTCCTATCAATGAGTGTATGATGCTCTCTATGCAGAGAAATCCTCAGATATTTCATCGC 1686
||| 61 ValSerIleAsnGluCysTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArg 80
QY 1687 CAACCTGGATTTCCTTTTAGTCGCAACGACAGTTAGAGAACCTATCACATATT 1746
||| 81 GlnProGlyPheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIle 100
QY 1747 CTTGTGTGCAGGCCAAACGAAACAAAGCAAGCATGCTCTGAATTTCTTGAATCTGAAGAT 1806
```

Db 101 LeuValCysArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAsp 120
QY 1807 GGGGAAGTAGAACAGAAAGACATATAGTAGTGGCCACAAATGCTGTATTTCCATAGT 1866
Db 121 GlyGluValGluGlnGlnArgThrTyrSerSerGlyHisAsnArg***TyrPheHisSer 140
QY 1867 GATACCTGCTTACCTCTCCGTCACAAAGAAATGGAAGTAGATGGAAGATGAAAAGGAT 1926
Db 141 AsphThrCysLeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAsp 160
QY 1927 CCTGAATGGCTAGAGAAAAACCATTAACAATTCGAAGGTTT 1972
Db 161 ProGluThrLeuArg-***LysProLeuHisLysLeuLysSerPhe 175

RESULT 13

AAU16381
ID AAU16381 standard; protein; 278 AA.
XX
AC AAU16381;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1334.
XX
KW Human; immunosuppressive; antiarthritic; antiinflammatory; cytostatic;
KW cardiac; vasotropic; cerebroprotective; neurotropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmological; vulnery;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.

FN WO200155322-A2.

XX
PD 02-AUG-2001.XX
PF 17-JAN-2001; 2001WO-US001341.XX
XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205551P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217486P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234597P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-SEP-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.

PR 08-NOV-2000; 2000US-0245613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-488783/53.
 XX N-PSDB; AAS26368.
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX
 PS Claim 11; SEQ ID NO 1334; 980pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Alignment Scores:

Pred. No.: 6.76e-63 Length: 278
 Score: 733.00 Matches: 142
 Percent Similarity: 98.62% Conservative: 1
 Best Local Similarity: 97.93% Mismatches: 2
 Query Match: 9.53% Indels: 0
 DB: 4 Gaps: 0
 US-09-874-162A-7 (1-4409) x AAU16381 (1-278)
 QY 1 CCGCGCCGCGCTCGCAGAGCCGACACCGAGGGGGCTCTCGATGTAGCACCATGACAGGC 60
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db ProAlaArgArgSerGlnAsnArgHisGlnGlySerGlyCysSerThrMetThrGly 38
 QY 61 ATCGCGCGCGCTCTCTCTCTCCATACCTGCGGATTCGGGGCTCGGACTCCACTTC 120
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db IleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGlyLeuHisPhe 58
 QY 121 CCCACCTGGCGCGCTCATCGAGCAGCATCGAGGACAACACATCGATCAGATCCACGG 180
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db ProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThrAspProArg 78
 QY 181 GTTTTGAAGAAACAGAAATTACAGAGCCCAACCTATGTTGCCCTGAGTTACATAATAGA 240
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db ValLeuGluLysGlnGluLeuGlnGlnProThrTy-ValAlaLeuSerTyIleAsnArg 98
 QY 241 TTCATGACAGATCGTCCCGCGAGAGCAGAGTCCCTAAAGAAAGATTTCAGCCGAAG 300
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db PheMetThrAspAlaAlaArgArgGluGlnGluSerLeuLysLysIleGlnProLys 118
 QY 301 CTCGCGTGACTGTGTCAGCTCAGTGTCTCGAGGAATGTGTCCTCCACCCACGCCAC 360
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db LeuSerLeuThrLeuSerSerSerValSerArgGlyAsnValSerThrProProArgHis 138
 QY 361 AGCAGTGGAGCCTTACTCCCGCGTACCCCGGAGCCACCCATCACCCCTCTCTTATCCGC 420
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db SerSerGlySerLeuThr***ProValThrProIleThrProSerSerPheArg 158
 QY 421 AGCAGCAGCTCCGACA 435
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db SerSerThrProThr 163
 RESULT 14
 ABUS5450
 ID ABUS5450 standard; protein; 278 AA.
 XX AC ABUS5450;
 XX
 XX DT 18-MAR-2003 (first entry)
 XX DE Human novel polypeptide #537.
 XX KW Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX OS Homo sapiens.
 XX PN US2002132753-A1.
 XX PD 19-SEP-2002.
 XX PF 17-JAN-2001; 2001US-00764864.
 XX PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.

(HELI-) HELIX RES INST.
(REAS-) RES ASSOC BIO TECHNOLOGY.

Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
Seki N, Yoshikawa T, Otsuka M, Nagahazi K, Masuho Y;

WPI: 2003-450961/43.
N-PSDB; ADB62280.

New polynucleotides and polypeptides, useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy.

Claim 1; Page: 223pp; English.

The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesising the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them, for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell regeneration. Membrane proteins, signal transduction-related proteins, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis, neurological diseases, cancer, tumors). The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a protein of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

Sequence 243 AA;
SQ

[illegible]

Search completed: August 25, 2004, 18:41:20
Job time : 461.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 25, 2004, 18:26:55 ; Search time 78 Seconds
(without alignments)
5836.382 Million cell updates/sec

Title: US-09-874-162A-7
Perfect score: 7690
Sequence: 1 cccgcggcgctcgagag.....tacataaagttcaataatat 4409

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 segs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xlp
-Q/cgn2_1/USFTO_spool_p/US09874162/runat 25082004 171845 2205/app query.fasta_1.4551
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09874162 @Cgn1.1.117 @runat 25082004 171845 2205 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -YGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUTS-COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1 | 238.5 | 3.1 | 604 | US-09-339-947A-8 | Sequence 8, Appli |
| 2 | 224 | 2.9 | 611 | US-09-339-947A-1 | Sequence 1, Appli |
| 3 | 159.5 | 2.1 | 912 | US-07-789-915A-8 | Sequence 8, Appli |
| 4 | 159.5 | 2.1 | 912 | US-08-005-002C-8 | Sequence 8, Appli |
| 5 | 159.5 | 2.1 | 912 | US-08-487-203A-8 | Sequence 8, Appli |
| 6 | 159 | 2.1 | 1494 | US-08-755-587-186 | Sequence 186, App |
| 7 | 156.5 | 2.0 | 3696 | US-09-134-001C-5080 | Sequence 5080, Ap |
| 8 | 154.5 | 2.0 | 1388 | US-09-572-191-2 | Sequence 2, Appli |
| 9 | 154.5 | 2.0 | 1388 | US-09-723-262-2 | Sequence 2, Appli |
| 10 | 154.5 | 2.0 | 1388 | US-09-723-219-2 | Sequence 2, Appli |
| 11 | 152.5 | 2.0 | 2482 | US-08-328-254-6 | Sequence 6, Appli |
| 12 | 150.5 | 2.0 | 1167 | US-09-803-671B-2 | Sequence 2, Appli |

| | | | | | | |
|----|-------|-----|------|---|-------------------|--------------------|
| 13 | 146.5 | 1.9 | 1235 | 1 | US-08-118-101A-2 | Sequence 2, Appli |
| 14 | 146 | 1.9 | 1972 | 4 | US-09-418-710-21 | Sequence 21, Appli |
| 15 | 143.5 | 1.9 | 7954 | 4 | US-09-417-485D-8 | Sequence 8, Appli |
| 16 | 142.5 | 1.9 | 2954 | 4 | US-09-150-867-1 | Sequence 1, Appli |
| 17 | 139 | 1.8 | 3248 | 1 | US-08-353-700-1 | Sequence 1, Appli |
| 18 | 139 | 1.8 | 3248 | 5 | PCT-US95-16216-1 | Sequence 1, Appli |
| 19 | 136 | 1.8 | 2184 | 4 | US-09-417-485D-6 | Sequence 6, Appli |
| 20 | 134 | 1.7 | 1388 | 2 | US-08-685-576-1 | Sequence 1, Appli |
| 21 | 134 | 1.7 | 2662 | 4 | US-09-595-684B-31 | Sequence 31, Appli |
| 22 | 132.5 | 1.7 | 1969 | 4 | US-09-418-710-72 | Sequence 72, Appli |
| 23 | 132.5 | 1.7 | 3418 | 2 | US-08-639-501-2 | Sequence 2, Appli |
| 24 | 132.5 | 1.7 | 3418 | 2 | US-08-603-753D-4 | Sequence 4, Appli |
| 25 | 132.5 | 1.7 | 3418 | 3 | US-09-044-946-2 | Sequence 2, Appli |
| 26 | 132.5 | 1.7 | 3418 | 3 | US-08-755-587-44 | Sequence 44, Appli |
| 27 | 132.5 | 1.7 | 3418 | 3 | US-09-044-908-2 | Sequence 2, Appli |
| 28 | 132.5 | 1.7 | 3418 | 3 | US-09-099-753-4 | Sequence 4, Appli |
| 29 | 132.5 | 1.7 | 3418 | 3 | US-08-986-106-4 | Sequence 4, Appli |
| 30 | 131 | 1.7 | 1061 | 4 | US-08-701-154A-5 | Sequence 5, Appli |
| 31 | 130.5 | 1.7 | 1036 | 3 | US-08-891-640-3 | Sequence 3, Appli |
| 32 | 130.5 | 1.7 | 2329 | 3 | US-08-755-587-16 | Sequence 16, Appli |
| 33 | 130 | 1.7 | 1032 | 4 | US-09-914-259-26 | Sequence 26, Appli |
| 34 | 128.5 | 1.7 | 907 | 3 | US-08-938-830-26 | Sequence 26, Appli |
| 35 | 128.5 | 1.7 | 907 | 3 | US-09-020-222-26 | Sequence 26, Appli |
| 36 | 128.5 | 1.7 | 1031 | 4 | US-09-914-259-24 | Sequence 24, Appli |
| 37 | 128 | 1.7 | 798 | 4 | US-09-203-453-5 | Sequence 5, Appli |
| 38 | 128 | 1.7 | 798 | 4 | US-09-900-236-5 | Sequence 5, Appli |
| 39 | 128 | 1.7 | 1164 | 4 | US-09-457-708-2 | Sequence 2, Appli |
| 40 | 128 | 1.7 | 1164 | 4 | US-09-950-046A-2 | Sequence 2, Appli |
| 41 | 128 | 1.7 | 1164 | 4 | US-09-976-594-989 | Sequence 989, App |
| 42 | 128 | 1.7 | 1388 | 2 | US-08-685-576-4 | Sequence 4, Appli |
| 43 | 128 | 1.7 | 2973 | 2 | US-08-821-355A-7 | Sequence 7, Appli |
| 44 | 128 | 1.7 | 2973 | 2 | US-09-003-687A-7 | Sequence 7, Appli |
| 45 | 128 | 1.7 | 2973 | 3 | US-09-136-605-7 | Sequence 7, Appli |

ALIGNMENTS

RESULT 1

; Sequence 8, Application US/09339947A
; Patent No. 6630616
; GENERAL INFORMATION:
; APPLICANT: YOSHIDA, No. 6630616umasa
; APPLICANT: YANAI, Yukihiko
; APPLICANT: KATO, Yoshihiro
; APPLICANT: HIRATSUKA, Junzo
; APPLICANT: TAKAHASHI, Shigeru
; APPLICANT: MIWA, Tatsushi
; TITLE OF INVENTION: GENE FOR FLORAL REGULATION AND METHODS FOR CONTROLLING
; TITLE OF INVENTION: FLOWERING
; FILE REFERENCE: 032735-001
; CURRENT APPLICATION NUMBER: US/09/339,947A
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: JP 1998-180065
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: JP 1999-179043
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: ZN FING
; LOCATION: (310)..(335)
US-09-339-947A-8

Alignment Scores:
Pred. No.: 4.7e-14 Length: 604
Score: 238.50 Matches: 130
Percent Similarity: 32.70% Conservative: 95
Best Local Similarity: 18.90% Mismatches: 242

| | | | |
|--|-------|--|------|
| Query Match: | 3.10% | Indels: | 221 |
| Db: | 4 | Gaps: | 22 |
| US-09-874-162a-7 (1-4409) x US-09-339-947A-8 (1-604) | | | |
| QY | 436 | GAGCAACACAGATCTATAGATTTCTTGGAACTCGGAATCTCATAGCACCAATATTTTG | 495 |
| Db | 27 | LySProValGluLeuTyAsnIlelleGlnArgSerIlelleAsnProAlaPheLeu | 46 |
| QY | 496 | CACAGACTCTTACTCATCTCAT-----CGAACTCCACAAACATC----- | 543 |
| Db | 47 | GlnArgCysLeuLeuTyIlelleHisAlaArgArgIysArgSerLeuIleThIle | 66 |
| QY | 544 | -----AAAAGAAACATTTAAAGTTGATGATGTATTATCAAAAGTAGAG | 588 |
| Db | 67 | SerLeuSerGlyGlyThrAsnLysGluLeuArgAlaGlnAsnIlePheProLeuTyVal | 86 |
| QY | 589 | AAAATGAAGAGAGCAAGATCTCATAGCTTGTCAGCTCAT----- | 630 |
| Db | 87 | LeuLeuAlaArgProThrAsnAsnValSerLeuGluGlyHisSerProIleTyArgPhe | 106 |
| QY | 631 | -----TTGCAGCTTACGTTTACTGTTTCTTCCACAAAATGATAGCCATCA | 678 |
| Db | 107 | SerArgAlaCysLeuLeuThrSerPheHisGluPheGlyAsnLysAspTyThrGluAla | 126 |
| QY | 679 | -----CCAACTCAGAAAT-----GAACAAATTTCTGTACCTCGGAAGTC | 720 |
| Db | 127 | ThrPheValIleProAspValLysAsnLeuAlaThrSerArgAlaCysSerLeuAsnIle | 146 |
| QY | 721 | CTGCTTGTGAAAGTTTGCACAAAAAGAAAG-----GATGAAGTTGTGCCAATA | 771 |
| Db | 147 | IleLeuIleSer-----CysGlyArgAlaGluGlnThrPheAspAsnAsnCysSerGly | 165 |
| QY | 772 | AGGCAAGTCCCAAGSPTAAAGACAG-----GTG | 801 |
| Db | 166 | AsnHisValGluGlySerThrLeuGlnLysLeuGluGlyLysCysPheTrpGlyLysIle | 185 |
| QY | 802 | CTTTGATCTCGACCTCAATCAACAAACCCGGAATTTCCCGTCCCTGCGAGTTTCC | 861 |
| Db | 186 | Profile-----AspLeuLeuAlaSerSerLeuGlyAsnCysValSerLeuSerLeu--- | 202 |
| QY | 862 | AGTAATGAATTTGAACCTAGTAACAGCCATATGTTGAAGTCTTACTCGTTGCTATTAGA | 921 |
| Db | 202 | ----- | 202 |
| QY | 922 | GTGACTCGTCAGGAGAGAGATTAAATGGAATGATTAATGAGAAACCAATGAAAT | 981 |
| Db | 203 | -----GlyHisThrValGluMet | 208 |
| QY | 982 | ATTGATGTCAATGAAGCTTCCAGCC-----AGAAAGAAACGAAATCGTGAGATGGG | 1035 |
| Db | 209 | SerSerThrValGluMetThrProSerPheLeuGluProLysPheLeuGluAspSerSer | 228 |
| QY | 1036 | GAAAGACATTGTTGCAAAATGACAGTATTGATATAAACAGCGCTTACAGCTTTTA | 1095 |
| Db | 229 | CysLeuThrPheCysSerGlnLysValAspAlaThrGlySerPheGlnLeuGlnValSer | 248 |
| QY | 1096 | GATGGGAATATGAAGTAGCCATGACGAAATGGAAGATCTCAATAAGCAAGAAAGA | 1155 |
| Db | 249 | IleSerAlaGlnGluAlaGlyAlaLysAspValSerGluSerProTySer----- | 265 |
| QY | 1156 | GCAACATGGGAGACTATTTGATGGGAAGGCTGCTCCATTCGAAACATTTTCTCAG | 1215 |
| Db | 266 | -----ValTySerTyAsnAspValPro----- | 274 |
| QY | 1216 | GGACCTACGTTGCAGTTTCACCTCTTCGTTGGACAGAGAGACCATGATAATCTACGGCT | 1275 |
| Db | 274 | ----- | 274 |
| QY | 1276 | CCTATTGCCAAACCTCTTCCCACTAGAAATTCAGAGAGTCTCCATCAGGAAACAAAGCCT | 1335 |
| Db | 274 | ----- | 274 |

```

US-09-339-947A-1
; Sequence 1, Application US/09339947A
; Patent No. 6630616
; GENERAL INFORMATION:
; APPLICANT: YOSHIDA, NO. 6630616umasa
; APPLICANT: YANAI, Yukihiko
; APPLICANT: KATO, Yoshihiro
; APPLICANT: HIRATSUKA, Junzo
; APPLICANT: TAKAHASHI, Shigeru
; APPLICANT: MIWA, Tatsushii
; TITLE OF INVENTION: GENE FOR FLORAL REGULATION AND METHODS FOR CONTROLLING
; FILE OF INVENTION: FLOWERING
; FILE REFERENCE: 032735-001
; CURRENT APPLICATION NUMBER: US/09/339,947A
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: JP 1998-180065
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: JP 1999-179043
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: ZN FING
; LOCATION: (306)..(327)
US-09-339-947A-1

Alignment Scores:
Pred. No.: 1,17e-12 Length: 611
Score: 224.00 Matches: 132
Percent Similarity: 33.82% Conservative: 102
Best Local Similarity: 19.08% Mismatches: 252
Query Match: 2.91% Indels: 206
Dbs: 4 Gaps: 23

US-09-874-162A-7 (1-4409) x US-09-339-947A-1 (1-611)
QY 436 GAGCCACACAGATCTATAGATTCTCGAAGTCGGAATCTCATAGCACCACATATTTTG 495
Db 27 LysProValGluLeuTyrAsnIleGlnArgAlaIleArgAsnProLeuPheLeu 46
QY 496 CACAGAACTCTTACTTACATGCTCATCGAACTCCAGAACACACATCAAAAGGAAACAA 555
Db 47 GlnA-GCysLeuHisTyrLysIleGluAlaLysHisLysArgArgIleGlnMetThrVal 66
QY 556 TTTAAAGTTGATGATGTTATCAAAAGTAGAGAAATCAAGGAGACCAAGAAATCTCAT 615
Db 67 PheLeu-----Sergly 70
QY 616 AGCTTGTGAGCTCATTTGAGCTTACGTTTACTGTTTCTTCCACAAAATGATAAGCCA 675
Db 71 AlaIleAspAlaGlyValGln----- 77
QY 676 TCACCAAACTCAGAAATGACAAATCTGTACCTCGGAAGTCCTGTTGAAAGTT 735
Db 78 -----ThrGlnLysLeuPheProLeuTyrIleLeuLeuAlaArgLeu 91
QY 736 TGCCCAAAAAAAGAAAGAGTAAAGTTGTCCAAATAGGCAA----- 777
Db 92 ValSerProLys-----ProValAlaGluTyrSerAlaValTyrArg 105
QY 778 -----GTTCCACAGGTAAAAAGCAGGTGCCT----- 804
Db 106 PheSerArgAlaCysIleLeuThrGlyGlyLeuGlyValAspGlyValSerGlnAlaGln 125
QY 805 -----TTGATTCCTGACCTCAAT-----CAACAAAACCCGGAAT--- 840
Db 126 AlaAsnPheLeuProAspMetAsnArgLeuAlaLeuGluAlaLysSerGlySerLeu 145
QY 841 -----TTCCCGCTCCCTTGCAGTTTCCAGTAATGAATTTGAACCTAGTAACAGGCCAT 891
146 AlaIleLeuPheIleSerPheAlaGlyAlaGlnAsnSerGlnPheGlyIleAspSerGly 165
892 ATGGTGAAGTCTTACTCGTTGCTTATTAGAGTGACTCGTCCAGGAAGAGAGATTTAAT 951
166 LysIleHisSer----- 169
952 GGAATGATTAATGAGAAACCAATGAAATATTGATGTCAATGAAGAGCTTCCAGCCAGA 1011
170 GlyAsnIleGlyGlyHisCys-----LeuTrpSerLysIleProLeuGln 184
1012 AGAAAACGAAATCGTAGGATGGGAAAAGACATTTGTTCACAAATGACAGTATTGAT 1071
185 -----SerLeuTyrAlaSerTrpGln 191
1072 AAAAACAGGCGCTTACAGCTTTTAGATGGGAATATGAATGACCATGAGGAAATGAA 1131
192 LysSerProAsnMetAspLeuGlyGlnArgValAspThrValSerLeuValGluMetGln 211
1132 GAATGTCCAATAAGCAAGAAAGACCAACATGGAGACTATTCTTGTGGAAGAGCGTG 1191
212 ProCysPheIleLysLeuLysSerMetSerGluGluLysCysValSer----- 227
1192 CCTCATTGCAAAACATTTCTCAGGAGCCTACGTTGAGTTCATCTTCTGTCGACAGGA 1251
228 -----IleGlnValProSerAsnProLeuThr----- 236
1252 GAGACCAATGATAAATCTACGGCTCTTATGTCACAAACCTTTCGCCACTAGAAATTCAG 1311
237 -----SerSerProGlnGlnValGlnValThrIleSerAlaGluGlu 251
1312 AGTCTCCATCAGGAAACAGCCTGTTTCAAGTTTAAACCTACTCAAACTATTGCTGTAAA 1371
252 ValGlySerThrGluLysSerProTyrSerPheSerTyrAsnAspIleSer---Ser 270
1372 GAATCATGTACTACAGATCTCAAAACAGAAAGAAAGATACTCCAATGAAACCGGA 1431
271 SerSerLeuLeuGlnIleIleArgLeuArg-----ThrGlyAsn----- 283
1432 CAAAAATTAAGAATATTTATCAGTTTCTCTATAACAACAATCAAGGCAACAACTGAA 1491
284 -----ValValPheAsnTyrArgTyrTyrAsnAsnLysLeuGlnLysThrGlu 299
1492 GCAAGAGATGACCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551
300 ValThrGluAspPheSerCysProPheCysLeuValLysCysAlaSerPheLysGlyLeu 319
1552 CTCAGCATCTTAAACTCTGCCATAGCAGATTTATCTTCAACTATCTTATCATCCAAA 1611
320 ArgTyrHisLeuProSerThrHisAspLeuLeuAsnPheGluPheTrpValThrGluGlu 339
1612 GGTCTAGGATAGATGTTTCTCAATGAGTGTGTTATGATGGCTCCTATGAGGAAATCCT 1671
340 PheGlnAlaValAsnValSerLeuLysThrGluThrMetIleSerLysVal---AsnGlu 358
1672 CAGATATTTCAT---GCCAACCTGGATTGCTTTT----- 1704
359 AspAspValAspProLysGlnThrPhePhePheSerSerLysLysPheArgArgArg 378
1705 -----AGTCGCAACGAGCA----- 1719
379 ArgGlnLysSerGlnValArgSerSerArgGlnGlyProHisLeuGlyLeuGlyCysGlu 398
1720 ---GTTAAGAGAACCTATCACAATATTTCTGTCGAGGCCAAACGA----- 1767
399 ValIleAspLysThrAspAlaHisSerValArgSerGluLysSerArgIleProPro 418
1767 ----- 1767
419 GlyLysHisTyrGluArgIleGlyGlyAlaGluSerGlyGlnArgValProProGlyThr 438
1768 -----ACAAAAGCA 1776

```

Db 439 SerProAlaAaspValGlnSerCysGlyAaspProAaspTyrValGlnSerIleAlaGlySer 458
 QY 1777 AGCATGCTGAATTCCTTGAATCTGAAGATGGGAGTAGACAGCAA-----AGAACAA 1830
 Db 459 ThrMetLeuGlnPheAlaLysThrArgLysIleSerIleGluArgSerAaspLeuArgAsn 478
 QY 1831 TATAGTAGTGGCCAAATGCTGTATTTCATATAGTAGTATACCTCTTACCTCTCCGTCCTCA 1890
 Db 479 ArgSerLeuLeuGlnLysArgGlnPhePheHisSerHisArgAlaGlnProMetAlaLeu 498
 QY 1891 CAAGAAATG-----GAAGTAGATAGTAGAATGAAAGGATCTGAAATGGCTTAAGAGAA 1944
 Db 499 GluGlnValLeuSerAaspArgAaspSerGluAaspGluValAaspAaspValAlaAaspPhe 518
 QY 1945 AAAACCATACACAAATGAGAGTTTCTCATGTATGATGAAGAGAGAGAGAGAGAGAGATG 2004
 Db 519 GluAaspArgAaspMetLeuAaspPheValAaspValThrLysAaspGluLysGlnMetMet 538
 QY 2005 AAACCTCGAATCTCCATGTCATGAAGCATGGGTTTATGTCGACAAATCAATGAATCAT 2064
 Db 539 HisMetTrpAsnSerPheValArgLysGlnArgValLeuAlaAaspGlyHisIleProTip 558
 QY 2065 GCCTGTATGCTGTTTCTAGAAAATATGACAGAGAAATTAAG---AAGAAATTTATGT 2121
 Db 559 AlaCysGluAlaPheSerArgLeuHisGlyProIleMetValArgThrProHisLeuIle 578
 QY 2122 CGAAACTTCATGCTTCATCTAGTCAGCATGTCATGCTTAACTTATATAGCATAATGTCA 2181
 Db 579 TrpCysTrpArgValPheMetValLysLeuTrpAsnHisGlyLeuLeuAaspAlaArgThr 598
 QY 2182 ATAGATAAAGCTGTATACCAAGCTCGTGAATGTCAG 2217
 Db 599 MetAsnAaspCysAsnThrPheLeuGlnLeuGln 610

RESULT 3

US-07-789-915A-8
 ; Sequence 8, Application US/07789915A
 ; Patent No. 5212058
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Rohan T.
 ; APPLICANT: Tobias, John W.
 ; APPLICANT: Varshavsky, Alexander
 ; TITLE OF INVENTION: Ubiquitin-Specific Proteases
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 ; STREET: Two Militia Drive
 ; CITY: Lexington
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02173
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: 19911108
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brook, David E.
 ; REGISTRATION NUMBER: 22,592
 ; REFERENCE/DOCKET NUMBER: MIT-5091AA
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-861-6240
 ; TELEFAX: 617-861-9540
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 912 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

US-07-789-915A-8
 Alignment Scores: Length: 912
 Pred. No.: 2,278-06 Matches: 154
 Score: 159.50 Conservative: 130
 Percent Similarity: 33.14% Mismatches: 324
 Best Local Similarity: 17.97% Indels: 249
 Query Match: 2.07% Gaps: 31
 DB: 1

US-09-874-162A-7 (1-4409) x US-07-789-915A-8 (1-912)

QY 20 GCCGACACCCAGGGGCTCTCGATAGCATGACAGAGGATCGCCGCCCTCTCTCT 79
 Db 6 AlaAsnLysGluGluSerTyr-SerMetTyrProLysThrSerSerProProProThr 25
 QY 80 TCTCCAATACCTGCCGATTCGGGGCTGGGACTCCACTTCCCACTCCGCGACCTCA 139
 Db 26 -----ProThrAsn 28
 QY 140 TCGAGCACATCGAGGACCAACACATCGATACAGATCCACGGGTTTGAAGAAACAAAGAA 199
 Db 29 MetGlnIleProIleTyrGlnAlaProLeuGlnMetTyrGlyTyr--ThrGlnAlaProT 48
 QY 200 TACAGAGCCAACTATGTTGCCCTGAGTTACATAAATAGATTATGACAGATGCTGCC 259
 Db 48 yzLeuTyrProThrGlnIleProAlaTyrSerPheAsnMetValAsnGlnAsnGlnProI 68
 QY 260 GCCGAGAGCAGGAGTCCCTAAAGAAAGATTCACGCGAAGCTCTCGCTGACTCTGTCCA 319
 Db 68 leTyrHisGlnSerGlySerProHisHisLeuProProGlnAsnAsnIleAsnGlyGlyS 88
 QY 320 GTCAGTGTCTCGAGGAATGTGCCACTCTCCCGACGCCACAGCAGTGGAGCCTTACTC 379
 Db 88 erThrThrAsnAsnAsnIleAsnLysLysLysLysLysLysLysLysLysLysLysLys 108
 QY 380 CCCCCTGACCCCATCACCCT 439
 Db 108 snAsnGlySerSerGlyAsnGlnGlyAlaAsnSerSerGlySerGlyMetSerTyrAsnL 128
 QY 440 CAACACAGATCTATAGATTCTTCGAACTCGGAATCTCATGACACCAATATTTTGCACA 499
 Db 128 ysSerHisThrTyrHisHisAsnTyrSerAsnAsnHisIle-----141
 QY 500 GAACCTTACTTACATGTCTCATCAAACTCCAGAACAAACATCAAAAGGAAACATTTA 559
 Db 142 -----ProMetMetAlaSerProAsnSerGlySerAsn-----152
 QY 560 AAGTTGATGATATGTTATCAAAAGTAGAGAAATGAAAGGAGAGCAAGAATCTCATAGCT 619
 Db 153 -----AlaGlyMetLysLysGlnThrAsnSerSerAsnGlyAsnGlyS 167
 QY 620 TGTAGCTCATTTGCGAGCTTACGTTTCTGTTTCTTCCACAAAATATGATAGCCATCAC 679
 Db 167 erSerAla-----ThrSerProSerT 174
 QY 680 CAAACTCAGAAAATGAACAAAATCTGTACCTCGAAGTCTGCTGTGAAGTTTGC 739
 Db 174 yzSerSerTyrAsnSerSerSerGlnTyrAsp-----LeuTyrLysPheAspV 190
 QY 740 ACAAAAAGAAAGATGTAGTTTCCAAATAGCAAGTCCCAAGGTAAAGAGAGG 799
 Db 190 alThrLysLeuLysAsn-----LeuLysGluAsnSerSerAsnLeuIleGlnL 206
 QY 800 TGCCTTTGATTCCTGACCTCATCAACAAACCCCGGA-----837
 Db 206 euProLeuPhe-----IleAsnThrThrGluAlaGluPheAlaAlaAlaSerValGlnA 224
 QY 838 -----AATTTCCCGTCCCTTGCGATTTCCAGTAATGAATTTGAACCTAGTAAAC- 885
 Db 224 rgTyrGluLeuAsnMetLysAlaLeuAsnLeuAsnSerGluSerLeuGlnAsnSerS 244
 QY 886 -----AGCCATATGTTGAGTCTTACTCGTTGCTATTAGAGTGACTC 928

| | | |
|------|--|------|
| 541 | spalalysIysSerSerIysSerIysSerPheGlnHisCysAlaIthrAlaAspIavall | 561 |
| 1886 | GTCCACAGAAATGGAGTAGATAGTAGAGAGATGAAAGGATCTCGAATGCTGAAGAGAA | 1945 |
| 561 | ysProAspGIuPhe-----TyrL-567 | |
| 1946 | AAACCATCACAAAATTGAAGAGTTTCTCATGCTTAATGAAGAGAGAAAGAAGTGATGA | 2005 |
| 567 | ysThrIeuSerThrIleProIysPheIysAspIeuGlnTrpGlyHisGlnGluAspAlag | 587 |
| 2006 | AATCTGGAAATCTCCATGTCATGACGAT-----GGGT | 2038 |
| 587 | IuGluPheLeuThrHisLeuLeuAspGlnLeuHisGluGluLeuIleSerAlaIleAspG | 607 |
| 2039 | TTATTGCTGACAAATCAAATCAATCATGCGCTGTATGCTTTTGTGAAGAAATATTGGACAGA | 2098 |
| 607 | lyIeuThrAspAsnGlu-----IleGlnAsnMetLeuGlnS | 619 |
| 2099 | AAATAATTAGAAGAAATTATGTGCGAAATTCATGCTTCATCTAGTCAGCATGCAT----- | 2154 |
| 619 | erIleAsnAspGIuGlnLeu---LysValPhePheIleArgAsnLeuSerArgTyrGlyL | 638 |
| 2155 | ----GACTTTTAATCTTATTAGCATAATGTCATAGATAAAGCTGTTTACCAAGTCGCGTG | 2209 |
| 638 | ysAlaGIuPhe-----IleIysAsnAlaSerProArgIeuIysG | 651 |
| 2210 | AAATGCGACAAAATTAGAAAAGGGGAATCTGCTTCCCTCGCAACGAAATAACTG | 2269 |
| 651 | IuLeuIleGIuLysTyr-----GlyValIleAsnAspAspSerThrG | 665 |
| 2270 | AAGACAAAATGGACAGCAAAATGGATTTTAGTCAAAATTAATCTCAAAAGAGAGACTTTGG | 2329 |
| 665 | IuGlu-----AsnGlyTrpHisGlu----- | 671 |
| 2330 | AAACAGATAGTGTCTCGGGGTTTCAAAACAGACAGCAAAAAACAAAA | 2376 |
| 672 | -----ValSerGlySerSerIysArgGlyLysIysThrLys | 683 |

US-08-005-002C-8
 : Sequence 8, Application US/08005002C
 : Patent No. 5494818
 : GENERAL INFORMATION:
 : APPLICANT: Baker, Rohan T.
 : APPLICANT: Tobias, John W.
 : APPLICANT: Varshavsky, Alexander
 : TITLE OF INVENTION: Ubiquitin-Specific Proteases
 : NUMBER OF SEQUENCES: 9
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Kevin W. Farrell, P.C.
 : STREET: P.O. Box 999
 : CITY: York Harbor
 : STATE: Maine
 : COUNTRY: U.S.A.
 : ZIP: 03911
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/005,002C
 : FILING DATE: 15-JAN-1993
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/789,915
 : FILING DATE: 08-NOV-1991
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Farrell, Kevin M.
 : REGISTRATION NUMBER: 35,505
 : REFERENCE/DOCKET NUMBER: MIT-5091AAZ
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 207-363-0058

TELEFAX: 207-363-0528
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 912 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-005-002C-8

Alignment Scores:

Pred. No.: 2,276-06 Length: 912
 Score: 159.50 Matches: 154
 Percent Similarity: 33.14% Conservative: 130
 Best Local Similarity: 17.97% Mismatches: 324
 Query Match: 2.07% Indels: 249
 DB: 1 Gaps: 31

US-09-874-162A-7 (1-4409) x US-08-005-002C-8 (1-912)

```

QY 20 GCCGACACAGGGGGCTCTCGATGACACATGACAGGATCGCGCGCGCTCTCTCT 79
DB 6 AlaAsnLysGluLysSerTyrSerMetTyrProLysThrSerSerProProThr 25
QY 80 TCTCCAAATACCTGCGATTCGGGGGCTCGGACTCCACTCCCACTCCCGCGACTCA 139
DB 26 -----ProThrAsn 28
QY 140 TCGACACATCGAGGACAAACACATCGATACAGATCCAGCGGTTTATAGAAAACAAGAT 199
DB 29 MetGlnIleProIleTyrGluAlaProLeuGlnMetTyrGlyTyr--ThrGlnAlaPro 48
QY 200 TACAGCAGCAACCTATGTTCCCTGAGTTACATAAATAGATTACATGACAGATCCTCCC 259
DB 48 yrLeuTyrProThrGlnIleProAlaTyrSerPheAsnMetValAsnGlnAsnGlnPro 68
QY 260 GCCGAGACAGGAGTCCCTAAGAGAGATTCAGCGGAGCTCTCGCTGACTCTGTCCA 319
DB 68 leTyrHisGlnSerGlySerProHisHisLeuProGlnAsnAsnIleAsnGlyGly 88
QY 320 GCTCAGTGTCTCGAGGAAATGTCCTCACTCCCGCCAGCCAGCAGTGGAGCGCTTACTC 379
DB 88 erThrThrAsnAsnAsnIleAsnLysLysLysTyrPheHisAsnGlyIleThrAsn 108
QY 380 CCCCGTGACCCCAATCCCGCTCTCTTCAATCCCGAGCAGCAGCTCCGACAGAGC 439
DB 108 snAsnGlySerSerGlyAsnGlnGlyAlaAsnSerSerGlySerGlyMetSerTyrAsnL 128
QY 440 CAACACAGATCTATAGATTTCTCGAATCGGAATCTCATAGACCAATATTTTGCACA 499
DB 128 ysSerHisThrTyrHisHisAsnTyrSerAsnAsnHisIle----- 141
QY 500 GAACCTCTTACTATGCTCTCATCGAACTCCAGAACACATCAAAAGAAACATTTA 559
DB 142 -----ProMetMetAlaSerProAsnSerSerGlySerAsn----- 152
QY 560 AAGTTGATGATGTTATCAAAAGTAGAGAAATGAAAGGAGACCAAGATCTCATAGCT 619
DB 153 -----AlaGlyMetLysLysGlnThrAsnSerSerAsnGlyAsnGly 167
QY 620 TGTGAGCTCATTTGCGAGCTACGTTTACTGCTTTCTTCCAAAATATGTAAGCCATCAC 679
DB 167 erSerAla-----ThrSerProSer 174
QY 680 CAACCTCAGAAATGAACAAATTTCTGTACCTCGAAGTCTGCTGTGAAAGTTGCC 739
DB 174 yrSerTyrAsnSerSerGlnTyrAsp-----LeuTyrLysPheAsp 190
QY 740 ACAAAAAGAAAGATGTAAGTTGTCCAAATAGGCAAGTTCCCAAGGTAAGAAAGCAGG 799
DB 190 alThrLysLeuLysAsn-----LeuLysGluAsnSerSerAsnIleGlnL 206
QY 800 TGCCTTTGATTCGACCTCAATCAACAAACCCGGA----- 837
  
```

```

DB 206 euProLeuPhe-----IleAsnThrThrGluAlaGluPheAlaAlaAsnValGlnA 224
QY 838 -----AATTTCCCGTCCCTTGCAGTTTCCAGTAATGAATTTGAACCTAGTAAC- 885
DB 224 rgTyrGluLeuAsnMetLysAlaLeuAsnLeuAsnSerGluSerLeuGluAsnSerSrv 244
QY 886 -----AGCCATATGGTGAAGTCTTACTCGTTGCTATTATTAGAGTGACTC 928
DB 244 alGluLysSerSerAlaHisHisThrLysSerHisSerIle----- 258
QY 929 GTCCAGGAAGAAGAGAGTTAATGAATGAATTAATGGAGAAACCAAT---GAAATATTG 985
DB 259 -----ProLysHisAsnGluValLysThrGluThrHisGlyGluGluGluA 275
QY 986 ATGCTAATGAAGAGCTTCCAGCCAGAGAAACGAATCGTGAGGATGGGAAAGACAT 1045
DB 275 spAlaHisAspLysLysProHisAlaSerLysAspAlaHisGluLeuLysLysThrG 295
QY 1046 TTGTTGCACAAAATGACAGTATTATTGATAAAACAGCGCTTACAGCTTTTAGATGGGAAT 1105
DB 295 luValLysLys----- 298
QY 1106 ATGAAGTAGCCATGAGGAAATGGAAGATGTCCAAATAGCAAGAAAGACGACATGGG 1165
DB 299 --GluAspAlaLysGlnAspArgAsnGluLysValIleGlnGluProGlnAlaThrValL 318
QY 1166 AGACTATTCTTGATGGGAAGGCTGCTCCATTCGAAACATTTTCTCAGGAGACCTACGT 1225
DB 318 euProValValAspLysLysGlu-----ProGluLysSerValGluGluAsn----- 333
QY 1226 TGCAGTTCTACTCTTCGTTGGACAGAGAGACCAATGATAAATCTACGCTCTCTATTGCC- 1284
DB 334 -----ThrSerLysThrSerSerProSerProSerProAlaA 347
QY 1285 -----AAACTCTTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAACAAAGC 1333
DB 347 laLysSerTrpSerAlaIleAlaSerAspAlaIleLysSerArgGlnAlaSerAsnLysT 367
QY 1334 CTGGTTAGTTAACTACTCAACTATGCTTGTGTTTAAAGATCAATGACTACAGATCTAC 1393
DB 367 hrValSerGlySerMetValThrLysThrProIleSerGlyThrThrAlaGlyValSerS 387
QY 1394 AAACAAGAAAGAAAGGATCTCCAAATGAAACCGACAAATAAATAAGATAATTATTATC 1453
DB 387 erThrAsnMetAlaAlaIleThrIleGlyLysSerSerProLeu----- 402
QY 1454 AGTTTCTCTATAACAACAATAACAAGCAACAACTGAAGCAAGAGATGACCTGCATTGCC 1513
DB 403 -----LeuSerLysGlnProGlnLysLysAspLysLysValP 416
QY 1514 CTGTGTACTCTGAACTGCCCAACTTATAGTTTACTCAAGCATCTTAACTCTGCC 1573
DB 416 roProSerThrLysGlyLeuGluProLeuGlySerIleAla-----LeuArgMetCysP 434
QY 1574 ATAGCAGATTTTCTCACTATGTTTATCATCCAAAGGTGCTAGGATAGATTCTTCA 1633
DB 434 heAspProAspPheIleSerTyrValLeuArgAsnLysAspValGluAsnLysIleProV 454
QY 1634 TCAATGAGTGTATGATGGCTCTTATGAGAAATCTTCAGAGATATTATCATCCCAACCTG 1693
DB 454 alHisSerIleIle-----ProArgIleIleIleAsnArgAlaA 467
QY 1694 GATTTGCTTTTAGTCGCAAGGACCACTTAAAGAGAACACCTTATCACATATTCTTGCT 1753
DB 467 snIleCysPhe-----MetSerSerValLeuGlnValLeuLeuTyrC 481
QY 1754 GCAGGCCA-----AAACGAAACAAACAAAGCATG- 1782
DB 481 ysLysProPheIleAspValIleAsnValLeuSerThrArgAsnThrAsnSerArgValG 501
QY 1782 ----- 1782
DB 501 lyThrSerSerCysLysLeuLeuAspAlaCysLeuThrMetTyrLysGlnPheAspLysG 521
  
```



```

1783 QY -----TCTGAATTTCTGAATCTGAAGATGGG-----GAAAGTAGAACACGAAA 1825
521 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
luThrTyrGluLysLysPheLeuGluAsnAlaAspAlaGluLysThrThrGluSerA 541
1826 QY GAACATATAGTAGTGGCCACATCGTCTGTATTTCCATAGTATACCTGCTTACCTCTCC 1885
541 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
spAlaLysLysSerLysSerLysSerPheGlnHisCysAlaThrAlaAspAlaValL 561
1886 QY GTCCCAAGAAATGGAAGTAGATAGTGAAGATGAAAGGATCCTGAATGGCTAAGAGAAA 1945
561 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
ysProAspGluPhe-----TyrL 567
1946 QY AAACATATACAAATTAAGAGTTTCTGATGTTAATGAAGGAGAGAAAGAGTGATGA 2005
567 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
ysThrLeuSerThrIleProLysPheLysAspLeuGlnTrpGlyHisGlnGluAspAlaG 587
2006 QY AACTCTGGAATCTCCATGTCATGAAGCAT-----GGGT 2038
587 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
luGluPheLeuThrHisLeuLeuAspGlnLeuHisGluLeuLeuSerAlaIleAspG 607
2039 QY TTATTGCTGACAAATCAATGAATCATGCTGTATGCTGTTGTGAGAAATATGACAGA 2098
607 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
lyLeuThrAspAsnGlu-----IleGlnAsnMetLeuGlns 619
2099 QY AATAATTAAGAGATTTATGTCGAACTTCATGCTTCATCTAGTCAGCATGAT----- 2154
619 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
erIleAsnAspGluGlnLeu-----LysValPhePheIleArgAsnLeuSerArgTyrGlyL 638
2155 QY -----GACTTTAATCTTATTAGCATAAATGTCATAGATAAAGCTGTTACCAAGCTCCGCTG 2209
638 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
ysAlaGluPhe-----IleLysAsnAlaSerProArgLeuLysG 651
2210 QY AAATGACGACAAATTAAGAAAGGGGGAATCTGCTCCCTGCAACAAAGAAATTAAGTG 2269
651 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
luLeuIleGluLysTyr-----GlyValIleAsnAspAspSerThrG 665
2270 QY AAGAACAAATGGGACAGCAAAATGATTTAGTGAATTAACCTCAAAAGAGAAAGCTTTGG 2329
665 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
luGlu-----AsnGlyTrpHisGlu----- 671
2330 QY AAACAGATAGTCTCAGGGGTTTCAAAACAGAGACAAAACAAAAA 2376
672 Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
ValSerGlySerSerLysArgGlyLysLysThrLys 683

```

RESULT 5

```

US-08-487-203A-8
Sequence 8, Application US/08487203A
Patent No. 5683904
GENERAL INFORMATION:
APPLICANT: Baker, Rohan T.
APPLICANT: Tobias, John W.
APPLICANT: Varshavsky, Alexander
TITLE OF INVENTION: Ubiquitin-Specific Proteases
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Kevin M. Farrell, P.C.
STREET: P.O. Box 999
CITY: York Harbor
STATE: Maine
COUNTRY: U.S.A.
ZIP: 03911
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,203A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/005,002

```

```

FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Farrell, Kevin M.
REGISTRATION NUMBER: 35,505
REFERENCE/DOCKET NUMBER: MIT-5091A3Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 207-363-0558
TELEFAX: 207-363-0528
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 912 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-203A-8

Alignment Scores:
Pred No.: 2,27e-06 Length: 912
Score: 159.50 Matches: 154
Percent Similarity: 33.14% Conservative: 130
Best Local Similarity: 17.97% Mismatches: 324
Query Match: 2.07% Indels: 249
DB: 1 Gaps: 31

US-09-874-162A-7 (1-4409) x US-08-487-203A-8 (1-912)
QY 20 GCCGACACACGGGGGCTCTCGATGTAGACACCATGACAGCATCGCGCGCTCTCTCT 79
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6 AlaAsnLysGluGluSerTyrMetTyrProLysThrSerSerProProProThr 25
QY 80 TCTCCAATACCTGCGGATTCGGGGCTGGGACTCCACTTCCCACTCCCGCCGACCTCA 139
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 -----ProThrAsn 28
QY 140 TCGAGCACATCGAGGACCAACCATCGATACAGATCCACGGGTTTGAAGAAACAGAAT 199
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 MetGlnIleProIleTyrGlnAlaProLeuGlnMetTyrGlyTyr--ThrGlnAlaPro 48
QY 200 TACAGACGCAACTATGTTGCCCTGAGTTACATAATAGATTATCATGACAGATGCTGCC 259
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 yrLeuTyrProThrGlnIleProAlaTyrSerPheAsnMetValAsnGlnAsnGlnPro 68
QY 260 GCCGAGACGAGGATCCCTAAAGAAAGAAATTCAGCCGAGATCTCGCTGACTCTGTCCA 319
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 leTyrHisGlnSerGlySerProHisLeuProGlnAsnAsnIleAsnGlyGlyS 88
QY 320 GCTCAGTGTCTCGAGGAATGTCTCCACTCCCGCCACGCCACAGCAGCTTACTCTC 379
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 erThrThrAsnAsnAsnAsnIleAsnLysLysLysTrpHisSerAsnGlyIleThrAsn 108
QY 380 CCCCCTGACCCACCATCCCTCTCTTCATTCGCGCAGCAGCACTCCGACAGAGC 439
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 snAsnGlySerSerGlyAsnGlnGlyAlaAsnSerSerGlySerGlyMetSerTyrAsnL 128
QY 440 CAACACAGATCTATAGATTTCTTCGAACTCGGAATCTCATPAGCACCATAATTTTGCACA 499
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 ysSerHisThrTyrHisHisAsnTyrSerAsnAsnHisIle----- 141
QY 500 GAACTCTTACTACATGCTCATCGAAATCCAGAACAAACATCAAAAGGAAACATTTA 559
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 -----ProMetMetAlaSerProAsnSerGlySerAsn----- 152
QY 560 AAGTTGATGATATGTTATCAAAAGTAGAGAAATGAAGGAGAGCAAGAAATCTCATAGCT 619
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 -----AlaGlyMetLysLysGlnThrAsnSerSerAsnGlyAsnGlyS 167
QY 620 TGTGAGCTCATTTGCAGCTTACGTTTACTGGTTTCTTCCACAAAAATGATAGCCATCAC 679
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 erSerAla-----ThrSerProSerT 174
QY 680 CAAACTCAGAAATGAACAAAATTTGTTACCTCGAAGTCTGCTGTGTAAGATTGTC 739
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 yrSerSerTyrAsnSerSerSerGlnTyrAsp-----LeuTyrLysPheAspV 190

```



```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1494 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-755-587-186

```

```

Alignment Scores:
Pred. No.: 3,218-06 Length: 1494
Score: 159.00 Matches: 177
Percent Similarity: 33.92% Conservative: 151
Best Local Similarity: 18.30% Mismatches: 355
Query Match: 2.07% Indels: 284
DB: 3 Gaps: 46

```

US-09-874-162A-7 (1-4409) x US-08-755-587-186 (1-1494)

```

QY 118 TTCGCCACCTGGCGACCTCATCGAGCACATCGAGCAACCAACCATCGATACAGATCCA 177
Db 370 PheLysThrAlaSerAsnLysGluLeuLysSerGluAsnAsnLysArgLysSerLys 389
QY 178 CGGGTTTGTAGAAACACAGAAATACAGAGCAACCTATGTTGCC--CTGAGITACATA 234
Db 390 MetLeuPheLysAspLeuGluHisTyrProThrAsnLeuAlaCysLeuGluLeuVal 409
QY 235 AATAGATTTCATGACAGATGCTCCCGCGAGAGAGAGTCCCTAAAGAAAGATTTCAG 294
Db 410 AsnThrSerLeuGluSerGlnLysLysProSerLysSer-----HisAlaLeuAsp 427
QY 295 CGAAGCTCTCCCTGACTCTGTCAGCTCAGTGTCTCGAGGAATGTCTCCACTCCCCCA 354
Db 428 ProGlnLeuAsnLeu--IleSerGlyPheValGln-----AsnSerThrTyrValSer 444
QY 355 CGCCACAGCAGTGGAGGCTTACTCCCGCGTGACCCACCCACCCCTCTCTTCA 414
Db 445 AspSerGluSerGlyHisThrAlaProProThrLeuSerLeuLysGlnAspPheAspSer 464
QY 415 TTCGCGACGACGACTCCGACAGAGCAACACACAGATC-----450
Db 465 AsnArgAsnLeuThrProSerGlnLysAlaGluLeuThrGluLeuSerThrIleLeuGlu 484
QY 451 -----TATAGATTCTTCGAATCTTCATCGAATCTCATAGCACCAATATTTTG 495
Db 485 GluSerGlySerGlnPheGluPheThrGlnPheArgLys-----ProSerHisIle 501
QY 496 CACAGAACTCTTACTTACATGCTCTCATCGAACTCCAGAAC-----AACATCAAAAGG 549
Db 502 IleGlnLysAsnProPheGluMetProGluAsnGlnLeuThrIleLeuAsnSerThrSer 521
QY 550 AAAACATTAAAGTTGATGATG-----TTATCAAAAGTAGAG 588
Db 522 LysGluTrpLysAspAspLeuHisLeuThrThrAsnAlaProSerIleGlnValAsp 541
QY 589 AAAATGAAA-----GGAGAGCAAGAATCTCATAGCTTGTGCAGCT---627

```

```

Db 542 SerLysLysSerGluGlyIleIleGlyGlyGlnLysPheAlaCysLeuSerArgThr 561
QY 628 -----CAITTCGACGCTTACGTTTACTTGGT 651
Db 562 SerCysAsnArgSerAlaSerGlyTyrSerThrAspLysAsnGluValGluPheArgGly 581
QY 652 TTCTTCCAC---AAAATGATAGCCATCACCAAACTCAGAA-----690
Db 582 PheTyrSerAlaArgGlyThrLysLeuAsnValGlySerGluAlaLeuGlnLysAlaLys 601
QY 691 -----AATGAACAAAATTCCTGT-----708
Db 602 LysLeuPheSerAspLeuGluAsnIleAsnGluGluThrSerValGluValAspArgSer 621
QY 709 -----ACCTGGAAGTCTCTTCTGTGAAAGTTTCCACAAA 744
Db 622 PheSerSerLysTyrAsnAspSerValSerMetIleGlnIleGluAspCysAsnAsp 641
QY 745 AAAAGAAAGGATGTAAAGTTGTCCTCAATAAGCAAGTCCACACAGTAAAGAGGAGGCT 804
Db 642 LysAsnAsnGlu-----ProAsnAsnLysCysArgLeu---652
QY 805 TTGATTCCTGACCTCAATCAACAAACCCCGAAATTTCCCGTCCCTTGCGAGTTTCCAGT 864
Db 653 -----IleLeuGln 655
QY 865 AATGAATTTGACACTAGTACGCCATATGTTGTAAGTCTTACTCGTGTCTATTAGN---921
Db 656 AsnAsnIleGluMetThrThrAspIlePheValGluGluTyrThrGluSerTyrArgArg 675
QY 922 -----GTGACTCGTCCAGGAAGAGAGAGATTTTAATGGAATG 957
Db 676 AsnThrGluAsnGluGlyAsnGlnCysThrAspAlaGlyArgAsnThrCysAsnSerGlu 695
QY 958 ATTAATGGA-----GAAACCAATGAAATATTGATGTCATGAAGAG-----999
Db 696 SerAspGlySerAspSerLysAsnAspThrValTyrIleHisGluGluAsnGly 715
QY 1000 CTTCCAGCCAGAAAGAAACGAAATCGTAGGATGGGAAAGACATTTGTTGCACAAATG 1059
Db 716 LeuProCysIleAspGlnHisAsnIleAsp-----LeuLysLeuPhe 729
QY 1060 ACAGTATTTGATAAAACAGCGCCTTACAGCTTTTAGATGGGAATATGAA-----1110
Db 730 SerGlnPheMetLysGluGlyAsnThrGlnIleLysGluGlyLeuSerAspLeuThrCys 749
QY 1111 GTAGCCATGACGAAATGGAAGATGTCCTCAATAGCAAGAAAGACCAACATGGGAGACT 1170
Db 750 LeuValMetLysAlaGluGluThrSerHisValThrMetSerAsnLysGlnGlnLeuThr 769
QY 1171 ATTCTTGATGGGAAGAGCTGCTCCCTCATTCGAAACATTT-----TCTCAGGACCTACG 1224
Db 770 AlaAsnThrGlyGlnAsnIleLysAspPheAspThrPheTyrLeuSerPheGlnThrAla 789
QY 1225 TTGAGTTCCTCTGTTGGACAGAGACCAACATGATAATCTACGGCTCTCTATTGCC 1284
Db 790 SerArgLysAsnIleArgValSerArgLysLeuAsnLysAlaArgSerLeuLeuAsn 809
QY 1285 AAACCTCTTCCCACTAGAAAT-----TCAGAGAGTCTCCATCAGGAAACCAAG 1332
Db 810 GlnLysTrpThrGluGluGluLeuAsnAsnPheSerAspSerLeuAsnSerGluLeuLeu 829
QY 1333 CTGTGT---TCAGTTAAACCTACTCAAACTATTGCTGTTTAAAGATCATTTGACTACAGAT 1389
Db 830 ProGlyIleAspIleLysLysThrAsp-----IleSerAsnHisGluValIleGlu 846
QY 1390 CTACAAACAGAAAGAAAGGATGACT-----1416
Db 847 AsnThrGluArgLysAspLysIleThrLysGluSerLeuIleGlyThrGluAsnIleLeu 866
QY 1417 -----CCAAATGAAACCCGAAAAATTAAGAATATTTTATCAGTTT 1458

```

Db 867 LeuIleLeuGlnGlnArgProGluSerLysLysIleLysLysIleLysGluSerAlaValLeu 886
QY 1459 CTCATAACAACAAACAGCAACAACTAAGCAAGAGATGACCTGCATTGCCCTTGG 1518
Db 887 GlyPheHisThrAlaSerGlyLysLysIleGluLeuThrLysGlu 901
QY 1519 TGTACTCTGAATCGCGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCTGCCATAGC 1578
Db 902 SerLeuAspLysValLysAsnLeuPheGluGluLysGluGlnAspAsnSer 918
QY 1579 AGATTATCTCAACTATGTTATCATCCAAAGGTGCTAGATAGATGTTCTTATCAAT 1638
Db 919 GluIleThrAsnPheSerHis-----ArgGlyAlaLysMetSerLysAspArgGlu 935
QY 1639 GAGTGTATGATGGC---TCCTATCGAGAAATCCTCAGGATATTTCATCGCAACCTCGGA 1695
Db 936 GluCysLysAspGlyArgGluLeuAlaCysGlyThrThrGluIleThrThrProGlu 955
QY 1696 TTTGCTTTTAGTCGCAAGCAGCAGTTAAGAGACACCTATCACATATTTCTGTGTGC 1755
Db 956 TyrGluGluThrHisSerSerLeuGluLysLysLysValSerAsnGluIleAlaAlaLeu 975
QY 1756 AGGCCA-----AAACGAACA----- 1770
Db 976 ArgProArgLeuLeuSerAspAsnLeuTyrLysGlnThrGluAsnLeuLysIleSerAsp 995
QY 1771 -----AAAGCAAGCATGCTCGAATTTCTTGAATCTGAAGTGGGAA----- 1812
Db 996 HisAlaSerGlnLysAlaValHisGluAsnThrGluLysGluThrAlaLysLysPro 1015
QY 1813 -----GTAGACACAAAGACATATAGTAGTGGCCACATCGT---CTGTATTTCCAT 1863
Db 1016 ThrMetTyrThrAsnGlnSerThrTyrSerAlaIleGluAsnSerProLeuThrPheThr 1035
QY 1864 AGTGATACCTGCTTACCTCTCCGCCCAAGAAATGGAAGTAGATAGTGAAGATGAA--- 1920
Db 1036 GlnAspThr-----GluLysLysPheSerValSerGluAlaSerLeu 1049
QY 1921 ---AAGGATCTGATGCTAAGAGAAACCACTTACAAATGAAGATTTTCTGAT 1977
Db 1050 PheGluAlaLysLysTrpLeuArgGlu-----GlyGluTrpAspAsp 1063
QY 1978 -----GTTAATGAAGGAGAGAAAGAGATGTAAGTCTGGAATCTCCATGTC 2025
Db 1064 GlnSerGluArgIleAsnAlaLysValAsnCysLysGluTyrProAspAspTyrVal 1083
QY 2026 -----ATGAAGCATGGTTTATGCTGACATCAATCAATGAATCAT 2064
Db 1084 GluAsnProSerCysGlyAsnSerSerAsnSerAlaIleThrGluAsnAspLysAsnHis 1103
QY 2065 GCCTGTATGCTGTTGTAGAAATATGACAGAAATAATTAAGAAGATTTATGTCGA 2124
Db 1104 -----LeuSerGluLysGlnGlySerThrTyrLeuSerAsnSerThrMetSer 1119
QY 2125 AAC---TTTCATGCTTCACTAGTCAGCATGTCATGAC-----TTTAATCTTATTAGC 2172
Db 1120 AsnSerTyrSerTyrHisProGlyPheCysHisSerSerGluValTyrAsnLysSerGlu 1139
QY 2173 ATAATGTC-----ATAGATAAGCTGTTACCAAGCTCGT 2208
Db 1140 TyrLeuSerArgSerLysIleAspAsnSerGlyIleGluProValIleLysAsnIleArg 1159
QY 2209 GAAATGCA----- 2217
Db 1160 GluArgLysAsnIleGlyPheSerGluIleMetSerProGlyArgGluAlaAspThrAsp 1179
QY 2218 -----CAAAAATTAGAAAGGGGGAATCTGCTTCCCT 2250
Db 1180 ProGlnSerValAsnAspIleCysValGluLysLeuAlaThrAsnSerSerCysLysAsn 1199
QY 2251 GCAAAAC----- 2262
Db 1200 LysAsnThrAlaIleLysValAlaIleSerAspSerAsnAsnPheAsnThrIleGlnLys 1219

RESULT 7

US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370

GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080

; LENGTH: 3696
; TYPE: PR1

; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Alignment Scores:

Pred. No.: 8,67e-06 Length: 3696
Score: 156.50 Matches: 140
Percent Similarity: 35.74% Conservative: 147
Best Local Similarity: 17.43% Mismatches: 327
Query Match: 2.04% Indels: 189
DB: 4 Gaps: 32

US-09-874-162A-7 (1-4409) x US-09-134-001C-5080 (1-3696)

QY 139 ATCGACACATCGAGACACCAACATCATGATCCACGGGTTTTAGAAAAACAGAA 198

Db 1926 LeuGluLysIleAsnAsnIleGlnProSerThrGlnThrLysThrAsnAlaLysGlnGlu 1945

QY 199 TTACAGCAGCCCAACCTATGTTGCCCTGAGTTACATAAAATAGATTTCATGACAGATCTGCC 258

Db 1946 IleAsnAspLysAlaGlnGluIleGlnIleAsnAsn---ThrProAspAlaThr 1964

QY 259 CCGCAGAGCAGAGTCCTCTAAAGAAAGATTCAGCCGAGCTCTCGTGAATCTGCTGCC 318

Db 1965 GluGluGluLysGlnAlaThrAsnArgValAsnAlaGlyLeuAlaGlnAlaIleGln 1984

QY 319 AGCTCAGTCTCGAGGAATGTCCTCCACTCCGCCACCCACGAGTGGAGCCCTTACT 378

Db 1985 -----AsnIleAsnAsn-----AlaHisSerThrGlnGluValAsn 1996

QY 379 CCCCCCGTAGCCCCCATCCCTCTCTTCAATTCGCGACGACGACCTCCGACAGAG 438

Db 1997 GluSerLysThrAsnSerIleAlaThrIleLysSerValGlnProAsnValIleLysLys 2016

QY 439 CCAACACAGATCTATAGATTTCTTCCGAATCTCGAATCTCATAGCACCATAATTTTGCAC 498

Db 2017 ProThrAlaIleAsnSerLeuThrGlnGluAlaAsn----- 2028

QY 499 AGAAGCTTACTTACTGTTCTCATCGAACTCCAGAACTCCAGAACTCCAGAACTCCAGAA 552

Db 2029 AsnGlnLysThrLeuIleGlyAsnAspGlyAsnAlaThrAspGluLysGluAlaAla 2048

QY 553 -----ACATTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGAAGGA 600
Db LysGlnLeuValThrGlnLysLeuAsnGluGlnIleGlnLysIleHisGlnSerThrGln 2068
QY 601 GAGCAAGATCTCATAGCTTGTCAGCTCAT-----TTCCAGCTTACGTTTACT 648
Db AspAsnGlnValAspAsnValLysAlaGlnAlaIleThrAlaIleLysLeuLeu 2086
QY 649 GGTGTTTCCCAAAAATGATAGCCATCACAAAATCAGAAAATGAACAAAATCTGTT 708
Db -----AsnAlaAsnAlaHisLysArgGlnAspAlaIle 2097
QY 709 ACCCTGGAAGTCTGCTGTTGAAAGTTTGCACAAAAGAAAGAGATGATGTTGTCGA 768
Db -----IleLeuThrAsnLeuAlaGluSerLysLysSerAspIleArgAla--- 2113
QY 769 ATAAGCAAGATCCACACAGTAAAAGCAGGTGCGCTTTGATCTCTGACCTCAATCAAA 828
Db -----AsnGlnAspAlaThrThrGluGluLysAsnThrAlaIleGlnSerIleAspThr 2132
QY 829 AAACCGGGAATTTCCGTCCTTCCGAGTTTCCAGTAAATGAAATTTGAACCTAGTAACAGC 888
Db -----LeuAlaGlnAlaAAsnAsnIleAsnGlnAlaAsnThr 2145
QY 889 CATATGGTGAAGTCTTACTCGTTGCTATTATTAGAGTGACTCGTCAGGAAGAGAGATT 948
Db -----GluAspGlyGlnLysLeu 2160
QY 949 AATGGAATGATTAAATGGA----- 966
Db GlnAlaLeuValLeuSerThrGlnThrLysThrGlnAlaLysAlaAspIleAlaGlnAla 2180
QY 967 -----GAAACCATGAAATATTGATGTCATGAAGAGAGTT 1002
Db IleGlyGlnArgSerThrIleAspGlnAsnGlnAsnAlaThrGluGluLysGln 2200
QY 1003 CCAGCCAGAAAGAAACGAATCGTCAG-----GATGGGAAAGACATTT 1047
Db GluAlaLeuGluArgLeuAsnGlnThrAsnGlyValAsnAspArgIleGlnAlaAla 2220
QY 1048 GTTGACAAATGACAGTATTTGAT-----AAAACAGCGCTTACAGCTTTTATGATGGGAA 1104
Db -----LeuAlaAsnGlnAsnValThrAspGluLysAsnAsnIleLeuGluThrIleArgAsnVal 2240
QY 1105 TATGAAGTAGCCATCAGCAAAATGAAGAATGCAATGAAGCAAGAAAGACACATGG 1164
Db GluProIleValIleValLysProLysAlaAsnGluIleArgLysAlaAlaGlu 2260
QY 1165 GAGACTATTCTTGATGGGAAGAGCTGCTCCATTCGAACATTTTCTCAGGACCTTACG 1224
Db GlnThrThrLeuIleAsnGln-----AsnGlnAspAlaThr 2272
QY 1225 TTG-----CAGTTCACTTCTGTTGGACAGAGAGACCAATGATAATCTACGGCT 1275
Db LeuGluGluLysGlnIleAlaLeuGlyLysLeuGluGluValLysAsnGluAlaLeuAsn 2292
QY 1276 CCAATTGCAAAACCTCTTGCCACTAGAAATTCAGAGAGTCTCCATCAGCAAAAC----- 1329
Db -----AlaHisSerAsnAspValLysIleAlaGluAsnAsnGly 2310
QY 1330 -----AGCCTGGTTCAGTTAAACCTACTCAACTATTGCT-----GTTAAAGATCA 1377
Db IleAlaLysIleSerGluValHisProGluThrIleIleLysArgAsnAlaLysGlnGlu 2330
QY 1378 TTGACTACAGATCTCAACAAGAAAGAAAGGATCTCCAAATGAAACCGACAAA 1437
Db -----GlnIleAspThrIleAsnAlaAsn----- 2345
QY 1438 TPAAGATATTTTATCAGTTTCTCTATACAAATACAGCAACAAACATGAAGCAAGA 1497
Db -----AsnLysSerThrAsnGluLysSerAlaAla 2356
QY 1498 GATGACCTGCATTGCGCTTGGTGTACTCTGAACCTGCCCAACATTTATGATTTACAG 1557

Db IleAsp-----ArgValAsnValAlaLysIleAspAlaIle----- 2368
QY 1558 CATCTTAAACTCTGCCATAGCAGATTTATCTTCAACTATCTTTATCATCCAAAAGGTGCT 1617
Db -----AsnAsnIleThrAsnAlaIleThrThr 2377
QY 1618 AGGATAGATGTTTCTATCAATGAGTGTATGATGGCTCTCTATCGAGAAATCTCTCAGAT 1677
Db -----ValAsn-----AspAlaLysAsnSerGlyAsnThrSerIle 2391
QY 1678 ATTCAATCGCAACTGATTTGCTTTTGTAGTCGCAACGAGCACAGTAAAGACACATCTC 1737
Db SerGlnIleLeuProSerThrAla-----ValLys----- 2401
QY 1738 ACACATATTTCTTGTGTCAGGCGCAAAACGAAACCAAGCATGCTGAATTTCTTGAA 1797
Db ThrAsnAlaLeuAlaAlaLeuAlaSerGluAlaLysAsnLysAsnAlaIleIleAspGln 2421
QY 1798 TCTGAGATGGGAAGTAGAACAGCAAGAACATATAGTAGTGGCCACAACTCTGCTGCTAT 1857
Db ThrProAsnAlaThrAlaGluGluLysGluGluAlaAsnAsnLysValAspArgLeu--- 2440
QY 1858 TTCCATAGTATGATCTGCTTACCTCTCCGTCACCAAGAAATGGAAGTAGATGAGAT 1917
Db -----GlnGlu 2442
QY 1918 GAAAGGATCCTGAATGGCTAAGAGAAAAAACCATTTACACAAAT----- 1962
Db GluAlaAspAlaAsnIleLeuLysAlaHisThrThrAspGluValAsnAsnIleLysAsn 2462
QY 1963 GAAGAGTTTCTGATGTTAATGAAGAGAGAAAGAGTGAAGAACTCTGGAATCTCCAT 2022
Db GlnAlaValGlnAsnIleAsnAlaValGlnValIleLysLysGlnAsnValLys 2482
QY 2023 GTCATGAAGCATGGTGTATTGCTGCAATCAATGAATCATGCTGCTGTTGTA 2082
Db AsnGlnLeuAsnGlnPheIle---AspAsnGlnLysLys-----IleIle 2496
QY 2083 GAAATTTATGGACAGAAATAATTAAGAAATTTATGTCGAACTTCATCTCATCTA 2142
Db GluAsnThrProAspAlaThrLeuGluGluLysAlaGluAlaAsnArgLeuLeuGlnAsn 2516
QY 2143 GTC-----AGCATCATGACTTTTAACTTATTATTAGCATATG 2178
Db ValLeuThrSerThrSerAspGluIleAlaAsnValAspHisAsnAsnGluValAspGln 2536
QY 2179 TCAATAGATAAGCTGTTACCAAGCTCCGTAATGCAGCAAAATAGAAAAAGGGGAA 2238
Db AlalaAspLysAlaArgProLysIleGluGluIleValProGlnValSerLysLysArg 2556
QY 2239 TCTGCTCCCTCGCAACGAGAAATACTGAAGACAAATCGGACAGCAAAATGGATTT 2298
Db AspValLeuAsnAlaIleGlnGluAlaPheAsnSerGlnThrGlnGluIleGlnGluAsn 2576
QY 2299 AGTCAATTAATCAACAAAGAGAAAGTGTGGAACAGATAGTCTCAGGGGTTCAAAA 2358
Db GlnGluAlaThrAsnGluGluLysThrGluAlaLeuAsnLysIleAsnGlnLeuAsn 2596
QY 2359 CAGAGCAAA 2367
Db GlnAlaLys 2599
RESULT 8
US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; their use

```

; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572.191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-572-191-2

Alignment Scores:
Pred. No.:      8,4e-06      Length:      1388
Score:          154.50      Matches:      184
Percent Similarity: 36.14%      Conservative: 129
Best Local Similarity: 21.25%      Mismatches: 306
Query Match:      2.01%      Indels:      248
DB:              4          Gaps:      40

US-09-874-162A-7 (1-4409) x US-09-572-191-2 (1-1388)
QY 169 ACAGATCCACGGTGTAAAG---AAACAAGATTACAGACGCCAACCTATGTCGCCTG 225
Db 169 ACAGATCCACGGTGTAAAG---AAACAAGATTACAGACGCCAACCTATGTCGCCTG 225
QY 609 ThrArgLysArgGlnLeuGluSerGluLeuGlnSerLeuGlnLysAlaAsnLeu 628
Db 609 ThrArgLysArgGlnLeuGluSerGluLeuGlnSerLeuGlnLysAlaAsnLeu 628
QY 226 AGTTACATAAATAGATTC---ATGACAGATGCTCCCGCCGAGAGCAGGAGTCCCTAAG 282
Db 226 AGTTACATAAATAGATTC---ATGACAGATGCTCCCGCCGAGAGCAGGAGTCCCTAAG 282
QY 629 AsnLeuGluAsnLeuGluAlaThrLysAlaCysLysArgGlnGluValSerGlnLeu 648
Db 629 AsnLeuGluAsnLeuGluAlaThrLysAlaCysLysArgGlnGluValSerGlnLeu 648
QY 283 AAGAGATTTCAGCCGAAGCTCTCGTCACTGCTCCAGCTCAGTGTCTCGAGGGAATGTG 342
Db 283 AAGAGATTTCAGCCGAAGCTCTCGTCACTGCTCCAGCTCAGTGTCTCGAGGGAATGTG 342
QY 649 AsnLysIleHisAlaGlu-----ThrLeuLysIle-----Ile 659
Db 649 AsnLysIleHisAlaGlu-----ThrLeuLysIle-----Ile 659
QY 343 TCCACTCCCCACGC-----CAGAGAGTGGAGCCTTACTCCCCCGTGAC 390
Db 343 TCCACTCCCCACGC-----CAGAGAGTGGAGCCTTACTCCCCCGTGAC 390
QY 660 ThrThrProThrLysAlaThrGlnLeuHisSerArg-----ProVal--- 673
Db 660 ThrThrProThrLysAlaThrGlnLeuHisSerArg-----ProVal--- 673
QY 391 CCACCATCACCCCTCC---TCTTCATTCCGACGACGACTCCGACAGAGCCACACAG 447
Db 391 CCACCATCACCCCTCC---TCTTCATTCCGACGACGACTCCGACAGAGCCACACAG 447
QY 674 ProLysLeuSerProGluMetGlySerPheGlySer----- 685
Db 674 ProLysLeuSerProGluMetGlySerPheGlySer----- 685
QY 448 ATCTATAGATTTCGAACTCGGAAT-----CTCATAGACCAATATTTTGCACAGA 501
Db 448 ATCTATAGATTTCGAACTCGGAAT-----CTCATAGACCAATATTTTGCACAGA 501
QY 686 -----LeuTyThrGlnAsnSerSerIleLeuAspAsnAspIleLeuAsnGlu 701
Db 686 -----LeuTyThrGlnAsnSerSerIleLeuAspAsnAspIleLeuAsnGlu 701
QY 502 ACTCTTACTTACATGCTCATCGAACTCCAGAACACACATCAAAAGGAAACATTTTAA 561
Db 502 ACTCTTACTTACATGCTCATCGAACTCCAGAACACACATCAAAAGGAAACATTTTAA 561
QY 702 ProValProProGluMetAsnGluGlnAlaPheGluAlaIleSerGluGluLeuArgThr 721
Db 702 ProValProProGluMetAsnGluGlnAlaPheGluAlaIleSerGluGluLeuArgThr 721
QY 562 GTTGATGATATGTTATCAAAAGTAGAGAAATGAAGAGAGCAGAAATCTCATAGCTTG 621
Db 562 GTTGATGATATGTTATCAAAAGTAGAGAAATGAAGAGAGCAGAAATCTCATAGCTTG 621
QY 722 ValGlnGluGlnMetSerAlaLeuGlnAlaLysLeuAspGluGluGlu----- 737
Db 722 ValGlnGluGlnMetSerAlaLeuGlnAlaLysLeuAspGluGluGlu----- 737
QY 622 TCAGCTCATTTGCAGCTTACGTTTACTGTTTCTTCCACAAAATGATAGCCATCACA 681
Db 622 TCAGCTCATTTGCAGCTTACGTTTACTGTTTCTTCCACAAAATGATAGCCATCACA 681
QY 738 -----HisLysAsnLeuLysLeuGlnGln 745
Db 738 -----HisLysAsnLeuLysLeuGlnGln 745
QY 682 AACTCAGAAAAT---GAACAAAATTCGTTTACCCTGGAAGTCCTGTTGGAAGTTTGC 738
Db 682 AACTCAGAAAAT---GAACAAAATTCGTTTACCCTGGAAGTCCTGTTGGAAGTTTGC 738
QY 746 HisValAspLysLeuGluHisSerThrGlnMetGlnGluLeuPheSerSer----- 763
Db 746 HisValAspLysLeuGluHisSerThrGlnMetGlnGluLeuPheSerSer----- 763
QY 739 CACAAAAAAGAGGATGTAAAGTTGTCCTCAATAGGCAAGTCCACAGGTAAGAACAG 798
Db 739 CACAAAAAAGAGGATGTAAAGTTGTCCTCAATAGGCAAGTCCACAGGTAAGAACAG 798
QY 764 -----GluArgIleAspThrLysGlnGln 772
Db 764 -----GluArgIleAspThrLysGlnGln 772
QY 799 GTGCTTTGATTCTCGACTCAAT-----CAACAAA 831
Db 799 GTGCTTTGATTCTCGACTCAAT-----CAACAAA 831
QY 773 GluGluLeuLeuSerGlnLeuAsnValLeuGluLysGlnLeuGlnThrGlnThrLys 792
Db 773 GluGluLeuLeuSerGlnLeuAsnValLeuGluLysGlnLeuGlnThrGlnThrLys 792
QY 832 CCCGGAATTTCCCGTCCCTTGCAGTTTCCAGTAAATGAATTTGAACCTAGTAACAGCAT 891
Db 832 CCCGGAATTTCCCGTCCCTTGCAGTTTCCAGTAAATGAATTTGAACCTAGTAACAGCAT 891
QY 793 -----AsnAspPheLeuLysSerGluValHis 801
Db 793 -----AsnAspPheLeuLysSerGluValHis 801

QY 892 -----ATGGTGAAGTCTTACTCG 909
Db 802 AspLeuArgValValLeuHisSerAlaAspLysGluLeuSerValLysLeuGluTyr 821
QY 910 TTGCTATTAGAGTGAAGTCCAGGAAAGAGAGAGATTTAATGGAATGATTAAATGGAGAA 969
Db 822 SerSerPheLysThrAsnGln-----GluLysGluPheAsnLysLeu----- 835
QY 970 ACCAATGAAATATTGATGCTCAATGAAGAGCTTCCAGCCAGAGAAAAACGAAATCGTGAG 1029
Db 836 SerGluArgHisMetHisValGlnLeuGlnLeuAspAsnLeuArgLeuGluAsnGluLys 855
QY 1030 GATGGGAAACACATTTTGTGCACAAATGACAGTATTGTATAAAACAGCGCTTACAG 1089
Db 856 LeuLeuGluSerLys--AlaCysLeuGlnAspSerTyAspAsnLeuGlnGluIleMetL 875
QY 1090 CTTTGTAGTGGGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1149
Db 875 ysPheGluIleAspGlnLeuSerArgAsnLeuGlnAsnPheLysLysGluAsnGluThrL 895
QY 1150 AAAAGAGCAACATGGAGACTATTCTTGTGGAAAGAGCTGCTCCATTTCGAAACATTT 1209
Db 895 euLysSerAspLeuAsnAsnLeuMetGluLeuGluAlaGluLysGluArgAsn--- 913
QY 1210 TCTCAGGACCTTACGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1269
Db 914 --AsnLysLeu--SerLeuGlnPheGlu-----GluAspLysGluAsnSer 927
QY 1270 ACGGTCTCTATTGCAAACTTCTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAAC 1329
Db 928 SerLysGluIleLeuLysValLeu-----GluAlaValArgGlnGluLys 942
QY 1330 AAGCCTGCTTCACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1377
Db 943 GlnLysGluThrAlaLysCysGluGlnGlnMetAlaLysValGlnLysLeuGluGluSer 962
QY 1378 TTGACTTACAGATCTACAAACAGAAAA-----GAAAGAGTACTTCCAAAT 1422
Db 963 Leu-----LeuAlaThrGluLysValIleSerSerLeuGluLysSerArgAspSer 979
QY 1423 GAAACACCGA-----CAAAATTAAGAAATATTTATCAG 1455
Db 980 AspLysLysValValAlaAspLeuMetAsnGlnIleGlnGluLeuArg----- 995
QY 1456 TTTCTCTATAACAACAATAACAGGCAACAACTGAAGCAAGAGAGTCACTGCTGCTCCT 1515
Db 996 -----SerSerValCysGluLysThrGluThrIleAspThrLeuLysGlnGlu 1011
QY 1516 TGGTGTACTCTGAATGCTGGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCTGCCAT 1575
Db 1012 LeuLysAspIleAsnCysLys-----Tyr--AsnSerAlaLe 1023
QY 1576 ACCAGATTATCTTCACTTATCATGTTTATCATCCAAAGGTGCTAGTAGATGTTCTTCTATC 1635
Db 1023 uValAspArgGluGluSerArgValLeuIleLys----- 1035
QY 1636 AATGAGTGTATGATGCTCTTATCCAGGAAATCTCAGGATATTCATCGCAACCTGGA 1695
Db 1036 -GlnGluValAspIleLeuAspLeuLysGluThrLeuArgLeuArgIleLeuSerGluAs 1055
QY 1696 TTTGCTTTTGTAGTCGCAACGGACCTTAAGAGACACCTATCACACATA----- 1744
Db 1055 p-----IleGluArgAspMetLeuCysGluAspLeuAlaHisAlaThrGluGlnLe 1072
QY 1745 -----TTCTTGTGTGACGCGCCAAAACGA 1767
Db 1072 uAsnMetLeuThrGluAlaSerLysLysHisSerGlyLeuLeuGlnSerAlaGlnGlu 1092
QY 1768 A-----CAAAGCAAGCATGCTCAATTTCTGAATCTGAAGATGG- 1809
Db 1092 uLeuThrLysLysGluAlaLeuIleGlnGluGlnHisLysLysAsnGlnLysLysG 1112
QY 1810 -GAAGTAGAACCAAGAAACA---TATAGTAGTGGCCACAATCGTCTG----- 1854

```

| | | | |
|----|------|---|------|
| Db | 1112 | uGluValGluGlnLysAsnGlnTyrAsnPheLysMetArgGlnLeuGluHisValMe | 1132 |
| Qy | 1855 | -----TATTTCATCATGATAC | 1871 |
| Db | 1132 | tAspSerAlaAlaGluAspProGlnSerProLysThrProPheHisGlnThrHisLe | 1152 |
| Qy | 1872 | CTGCTTACCTTCGCTCCACAAGAAATGGAGTAGATAGTGAAGATGAAGAAGATCCTGA | 1931 |
| Db | 1152 | uAlaLysLeuLeuGluThrGlnGluGlnGluLe-----GluAspGlyAlaSerLys | 1170 |
| Qy | 1932 | ATGCTAAGAGAAAAACCATTACAAAAATGAAGAGCTTTTCTCATGCTTAATGAAGAGA | 1991 |
| Db | 1170 | sThrSerLeuGluHisLeuValThrLysLeuAsnGluAspArgGluValLysAsnAlaGl | 1190 |
| Qy | 1992 | G-----AAAGAAGTGAATGAACCTCGGAATCTCATGCTCATGAAGCATGG | 2036 |
| Db | 1190 | uIleLeuArgMetLysGluGlnLeuArgGluMetGluAsnLeuArgLeuGluSerGlnGl | 1210 |
| Qy | 2037 | GTATTATTGCTGACAAAT----- | 2052 |
| Db | 1210 | nLeuIleGluLysAsnTrpLeuLeuGlnGlyGlnLeuAspAspIleLysArgGlnLysGl | 1230 |
| Qy | 2053 | -----CAATGAATCATGCTGTATGCTGTTGTAGAAAAATATTGGACAGAAAAATAAT | 2105 |
| Db | 1230 | uAsnSerAspGlnAsnHisProAspAsnGlnGlnLeuLysAsnGluGlnGluSerIl | 1250 |
| Qy | 2106 | TAAGAAGAATTTATGTGCAAACTTCATGCTTCAT-----CTAGTCAGCATGCGATGACTT | 2159 |
| Db | 1250 | eLysGluArgLeuAlaLysSerLysIleValGluGluMetLeuLysMetLysAlaAspLe | 1270 |
| Qy | 2160 | TAATCTTTATTAGCATATGTCAATAGATAAGCTGTTTACCAGCTCCGT----- | 2208 |
| Db | 1270 | uGluGluValGlnSerAlaLeuTyrAsnLysGluMetGluCysLeuArgMetThrAspGl | 1290 |
| Qy | 2209 | ----GAAATGCAGCAAAATTAGAA-----AAGGGGGAATCTGCTTCCCC | 2249 |
| Db | 1290 | uValGluArgThrGlnThrLeuGluSerLysAlaPheGlnGluLysGluGlnLeuArgSe | 1310 |
| Qy | 2250 | TGCAAACGACAGAATAACTAGNACAAAAATGGACAGCAAAATGGATTAGTGAAATTAA | 2309 |
| Db | 1310 | rLysLeuGluGluMetTyrGluGluArgGluA-gThrSerGlnGluMetGluMetLeuAr | 1330 |
| Qy | 2310 | CTCAAAAGAGAAAGCTTTGGAAAAACAGATAGTGTCTCAGGGGTTCAAAAACAGAGAAAAA | 2369 |
| Db | 1330 | gLysGlnValGluCysLeuAlaGluGluAsnGlyLysLeuValGlyHisGlnAsnLeuHi | 1350 |
| Qy | 2370 | ACAAAAACTC | 2379 |
| Db | 1350 | sGlnLysIle | 1353 |

RESULT 9

```

US-09-723-262-2
; Sequence 2, Application US/09723262
; Patent No. 6379912
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; APPLICANT: No. 6379912el motor proteins and methods for
; TITLE OF INVENTION: No. 6379912el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,262
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-723-262-2

```

Db 836 SerGluArgHisMetHisValGlnLeuGlnLeuAspAsnLeuArgLeuGluAsnGluLys 855
 QY 1030 GATGGGAAAGACATTTGTTGCAAAATGACAGATATTGTTATATAAAACAGCGCGTTACAG 1089
 Db 856 LeuLeuGluSerLys--AlaCysLeuGlnAspSerTyrAspAsnLeuGlnGluMetL 875
 QY 1090 CTTTATAGATGGGATATAGACATGACGAGGAATGGAAGATGTCACATGAAGCAAG 1149
 Db 875 ysPheGluLeuAspGlnLeuSerArgAsnLeuGlnAsnPhelysGluAsnGluThrL 895
 QY 1150 AAAAGAGCAACATGGAGGACTATTCTTGATGGAGAGGCTGCCTCCATTCGAAACATTT 1209
 Db 895 euLysSerAspLeuAsnAsnLeuMetGluLeuLeuGluAlaGluLysGluArgAsn--- 913
 QY 1210 TCTCAGGACCTACGTTCCAGTTCTACTCTTCTGTTGGACAGAGACCAATGATAATCT 1269
 Db 914 --AsnLysLeu--SerLeuGlnPheGlu-----GluAspLysGluAsnSer 927
 QY 1270 ACGGCTCTATTGCGCAACCTCTTGGCCACTAGAAATTGAGAGAGTCTCCATCAGAAAC 1329
 Db 928 SerLysGluLeuLeuLysValLeu-----GluAlaValArgGlnGluLys 942
 QY 1330 AAGCCTGTTGTTAACTTACTTCAACTATTGCT-----GTTAAAGATCA 1377
 Db 943 GlnLysGluThrAlaLysCysGluGlnGlnMetAlaLysValGlnLysLeuGluGluSer 962
 QY 1378 TTGACTACAGATCTACAAACAGAA-----GAAAGGATACTCCAAAT 1422
 Db 963 Leu-----LeuAlaThrGluLysValLysSerSerLeuGluLysSerArgAspSer 979
 QY 1423 GAAACCGA-----CAAAATTAAGATATTTATCAG 1455
 Db 980 AspLysLysValValAlaAspLeuMetAsnGlnLeuGlnLeuArg----- 995
 QY 1456 TTTCTCTATACAACTATACAGGCAACAACTGAGCAAGAGATGACCTGCATGCGCCT 1515
 Db 996 -----SerSerValCysGluLysThrGluThrLeuAspThrLeuLysGlnGlu 1011
 QY 1516 TGGTGTACTCTGAAGTGGCGCAAACTTTATGTTTACTAGCATCTTAAACTCTGCCAT 1575
 Db 1012 LeuLysAspLysAsnCysLys-----Tyr--AsnSerAlaLe 1023
 QY 1576 AGCAGATTATCTCAACTATGTTTATCATCAAAAGGTGCTAGATAGATGTTTCTATC 1635
 Db 1023 uValAspArgGluGluSerArgValLeuLysLys----- 1035
 QY 1636 AATGAGTCTTATGATGGCTCTATGAGGAATCTCAGGATATTCAGCACTGGA 1695
 Db 1036 -GlnGluValAspLysLeuAspLysGluThrLeuArgLeuArgLeuSerGluAs 1055
 QY 1696 TTTGCTTTTATGTCGCAACCGACCACTTAAGAGACACTATCACATA----- 1744
 Db 1055 p-----IleGluArgAspMetLeuCysGluAspLeuAlaHisAlaThrGluGlnLe 1072
 QY 1745 -----TTCTGTGTGCGCGCCAAACGA 1767
 Db 1072 uAsnMetLeuThrGluAlaSerLysLysHisSerGlyLeuLeuGlnSerAlaGlnGlu 1092
 QY 1768 A-----CAAAAGCAAGCATCTCTGAATTTCTTGAATCTGAAGATGGG----- 1809
 Db 1092 uLeuThrLysLysGluAlaLeuLeuGlnGlnHisLysLeuAsnGlnLysLysG 1112
 QY 1810 -GAAGTAGACAGCAAAACA-----TATAGTAGTGGCCACCAATCGTCTG----- 1854
 Db 1112 uGluValGlnGlnLysLysAsnGluTyrAsnPheLysMetArgGlnLeuGluHisValMe 1132
 QY 1855 -----TATTTCATGATGATAC 1871
 Db 1132 tAspSerAlaAlaGluAspProGlnSerProLysThrProHisPheGlnThrHisLe 1152
 QY 1872 CTGCTTACTCTCCGTCACAGAAATGGAAGTAGATGATGAAAGATCTCGA 1931
 Db 1152 uAlaLysLeuLeuGluThrGlnGlnGluLe-----GluAspGlyArgAlaSerLys 1170

QY 1932 ATGGCTAAGAGAAAAACCATTTACAAATTCAGAGATTTTCTGATGTTAATGAAGGAGA 1991
 Db 1170 sThrSerLeuGluHisLeuValThrLysLeuAsnGluAspArgGluValLysAsnAlaG 1190
 QY 1992 G-----AAAGAAGTGCATGAAATCTCGAATCTCCATGTCATGACAGCATGG 2036
 Db 1190 uLeuLeuArgMetLysGluGlnLeuArgGluMetGluAsnLeuArgLeuGluSerGlnG 1210
 QY 2037 GTTTATTGCTGCAAT----- 2052
 Db 1210 nLeuLeuGluLysAsnTTPLeuLeuGlnGlyGlnLeuAspAspLysLysArgGlnLysG 1230
 QY 2053 -----CAATGAATCATGCTGATGCTGTTGTAGAAAATTTGACACAGAAATAAT 2105
 Db 1230 uAsnSerAspGlnAsnHisProAspAsnGlnGlnLeuLysAsnGlnGluGluSerIl 1250
 QY 2106 TAAGAAGATTTATGTCGAAACTTCATCTTCAT-----CTAGTCAGCATGCGATGTT 2159
 Db 1250 elysGluArgLeuAlaLysSerLysLeuValGluGluMetLysMetLysAlaAspLe 1270
 QY 2160 TAACTTTATGACATAATGCTCAATAGATAAAGCTGTTACCAAGCTCCCT----- 2208
 Db 1270 uGluGluValGlnSerAlaLeuTyrAsnLysGluMetGluCysLeuArgMetThrAspG 1290
 QY 2209 -----GAAATGCAGCAAAAATTAGAA-----AAGGGGAATCTGCTTCCCC 2249
 Db 1290 uValGluArgThrGlnThrLeuGluSerLysAlaPheGlnGluLysGluGlnLeuArgSe 1310
 QY 2250 TGCAACAGAGAAATACTGAAGAACAAAATCGGACACAAATGGATTAGTGAATTA 2309
 Db 1310 rLysLeuGluGluMetTyrGluGluArgThrSerGlnGluMetGluMetLeuArg 1330
 QY 2310 CTCAAAAGAGAAAGCTTTGGAAACAGATAGTCTCAGGGGTTTCAAAACAGAGCAAAA 2369
 Db 1330 gLysGlnValGluCysLeuAlaGluGluAsnGlyLysLeuValGlyHisGlnAsnLeuHi 1350
 QY 2370 ACAAAACTC 2379
 Db 1350 sGlnLysIle 1353

RESULT 10

US-09-723-219-2
 ; Sequence 2, Application US/09723219
 ; Patent No. 6391613
 ; GENERAL INFORMATION:
 ; APPLICANT: Beraud, Christophe
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Wood, Kenneth
 ; TITLE OF INVENTION: No. 6391613el motor proteins and methods for
 ; FILE OF INVENTION: their use
 ; FILE REFERENCE: 1017
 ; CURRENT APPLICATION NUMBER: US/09/723,219
 ; CURRENT FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: US 09/572,191
 ; PRIOR FILING DATE: 2000-05-17
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 1388
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-723-219-2

Alignment Scores:
 Pred. No.: 8,4e-06 Length: 1388
 Score: 154.50 Matches: 184
 Percent Similarity: 36.14% Conservative: 129
 Best Local Similarity: 21.25% Mismatches: 306
 Query Match: 2,01% Indels: 248
 DB: 4 Gaps: 40

US-09-874-162A-7 (1-4409) x US-09-723-219-2 (1-1388)

| | | | |
|----|------|--|------|
| QY | 169 | ACAGATCCACCGGTTTTAGAA---AAACAAGAAATTACAGCAGCCAAACCTATGTTCCCTCG | 225 |
| DB | 609 | ThrArgLysArgGlnLeuGlnLeuGlnSerLeuGlnLysAlaAsnLeu | 628 |
| QY | 226 | AGTTACATAATAGATTCC---ATGACAGATGTCGCCCGCAGACAGAGAGTCCCTAAAG | 282 |
| DB | 629 | AsnLeuGluAsnLeuLeuGluAlaThrLysAlaCysLysArgGlnGluValSerGlnLeu | 648 |
| QY | 283 | AAGAAGATTACCGCAGAGCTCTCGCTGACTCTGTCCAGCTCAGTCTCTCGAGGGGAATGTG | 342 |
| DB | 649 | AsnLysIleHisAlaGlu---ThrLeuLysIle---:::-----Ile | 659 |
| QY | 343 | TCCACTCCCCCAGCG---CACACAGTGGAAAGCCTTACTCCCCCGCTGACC | 390 |
| DB | 660 | ThrThrProThrLysAlaThrGlnLeuHisSerArg---ProVal--- | 673 |
| QY | 391 | CCACCCATCACCCCTCC---TCTTCATTCGCGACAGCAGCACTCCGACAGAGCCACACAG | 447 |
| DB | 674 | ProLysLeuSerProGluMetGlySerPheGlySer---:::----- | 685 |
| QY | 448 | ATCTATAGATTCTTCGAACCTCGGAAT---CTCATAGCACCAATATTTTGCACAGA | 501 |
| DB | 686 | -----LeuThrGlnAsnSerSerIleLeuAspAsnAspIleLeuAsnGlu | 701 |
| QY | 502 | ACTCTTACTTACTGTCTCATCGAAACTCCAGAACAAACATCAAAAGAAACATTTAAA | 561 |
| DB | 702 | ProValProGluMetAsnGluGlnAlaPheGluAlaIleSerGluGluLeuArgThr | 721 |
| QY | 562 | GTTGATGATATGTTATCAAAAGTAGAGAAATGAAGAGAGCAGAAATCTCATAGCTTG | 621 |
| DB | 722 | ValGlnGluGlnMetSerAlaLeuGlnAlaLysLeuAspGluGluGlu---:::----- | 737 |
| QY | 622 | TCAGTCTATTTCGACCTTACGTTTACTGTTTCTTCCACAAAAATGATAAGCCATCACCA | 681 |
| DB | 738 | -----HisLysAsnLeuLysLeuGlnGln | 745 |
| QY | 682 | AACTCAGAAAT---GAACAAAAATCTGTTTACCTCGGAAGTCCTGTTGTGAAATTTGC | 738 |
| DB | 746 | HisValAspLysLeuGluHisSerThrGlnMetGlnGluLeuPheSerSer---:::----- | 763 |
| QY | 739 | CACAAAAAAGAAAGGATGTAAATTGTCCATAAGGCAAGTCCACAGAGTAAAGACGAC | 798 |
| DB | 764 | -----GluArgIleAspTrpThrLysGlnGln | 772 |
| QY | 799 | GTGCCTTTGATTCCTCGACCTCAAT---:::-----CAAAACAAA | 831 |
| DB | 773 | GluGluLeuLeuSerGlnLeuAsnValLeuGluLysGlnLeuGlnGluThrGlnThrLys | 792 |
| QY | 832 | CCCGGAAATTTCCGTCCTTCAGTTTCCAGTATGAAATTGAACCTAGTAACAGCCAT | 891 |
| DB | 793 | -----AsnAspPheLeuLysSerGluValHis | 801 |
| QY | 892 | -----ATGGTGCAAGCTCTTACTCG | 909 |
| DB | 802 | AspLeuArgValValLeuHisSerAlaAspLysGluLeuSerSerValLysLeuGluTyr | 821 |
| QY | 910 | TTGCTATTTAGAGTACTCGTCCAGGAAGAAGACAGAGTTTATGGAATGATTATATGAGAA | 969 |
| DB | 822 | SerSerPheLysThrAsnGln---GluLysGluPheAsnLysLeu---:::----- | 835 |
| QY | 970 | ACCAATGAAATATTTGATGTCAATGAAGAGCTTCAGCCAGACAAACGAAATCGTCAG | 1029 |
| DB | 836 | SerGluArgHisMetHisValGlnLeuLeuAspAsnLeuArgLeuGluAsnGluLys | 855 |
| QY | 1030 | GATGGGGAAGAAGCATTTGTTGCACAATGACAGATTTTGATAAAAAACAGCGCGTTACAG | 1089 |
| DB | 856 | LeuLeuGluSerLys--AlaCysLeuGlnAspSerTyrAspAsnLeuGlnGlnIleMetL | 875 |
| QY | 1090 | CTTTTAGATGGGAATATGAGTAGCCATGACGAGGAATGGAAGAAATCTCCAATAAGCAAG | 1149 |
| DB | 875 | ysPheGluIleAspGlnLeuSerArgAsnLeuGlnAsnPheLysGlnAsnGluThrL | 895 |

| | | | |
|----|------|--|------|
| QY | 1150 | AAAAGACACATGGGAGACTATCTTCGATGGGAGAGAGCTGCCTCATTCGAAACATTT | 1209 |
| Db | 895 | eulysSerAspLeuAsnAsnLeuMetGluLeuLeuGluAlaGluLysGluArgAsn--- | 913 |
| QY | 1210 | TCTCAGGCACTACGTTGCAGTTCACCTCTTCGTTGGACAGAGACCAATGATAAATCT | 1269 |
| Db | 914 | --AsnLysLeu--SerLeuGlnPheGlu-----GluAspLysGluAsnSer | 927 |
| QY | 1270 | ACGGCTCCTATTGCGAAACCTCTTGCCCATAGAAATTCAGAGATCTCCATCAGGAAAC | 1329 |
| Db | 928 | SerLysGluIleLeuLysValLeu-----GluAlaValArgGlnGluLys | 942 |
| QY | 1330 | AAGCCTGGTTCAGTTAAACCTACTCCTAACTATTGCT-----GTTAAAGATCA | 1377 |
| Db | 943 | GlnLysGluThrAlaLysCysGluGlnGlnMetAlaLysValGlnLysLeuGluGluSer | 962 |
| QY | 1378 | TTGACTACAGATCTACAAACAAGAAA-----GAAAGGATACTCCAAAT | 1422 |
| Db | 963 | Leu-----LeuAlaThrGluLysValIleSerSerLeuGluLysSerArgAspSer | 979 |
| QY | 1423 | GAAAACCGA-----CAAAANTTAAGATATTTATCAG | 1455 |
| Db | 980 | AspLysLysValValAlaAspLeuMetAsnGlnIleGlnGluLeuArg----- | 995 |
| QY | 1456 | TTTCTCTATACCAACATACAAAGCAAACTGAAGCAAGACAGATACCTGCATTGCCCT | 1515 |
| Db | 996 | -----SerSerValCysGluLysThrGluThrIleAspThrLeuLysGlnGlu | 1011 |
| QY | 1516 | TGGTGACTCTGAACGTGCCCAACTTTATAGTTTACTCAAGCATCTTAACTCTGCCAT | 1575 |
| Db | 1012 | LeuLysAspIleAsnCysLys-----TyrAsnSerAlaLe | 1023 |
| QY | 1576 | AGCAGATTTATCTCAACTATGTTTATCATCCAAAAGGTGCTAGGATAGATGTTCTATC | 1635 |
| Db | 1023 | uValAspArgGluGluSerArgValLeuIleLysLys----- | 1035 |
| QY | 1636 | AATGAGTGTTATGATGCTCCTATGAGGAAATCTCTCAGGATATTCATCCCAACCTGGA | 1695 |
| Db | 1036 | -GlnGluValAspIleLeuAspLeuLysGluThrLeuArgLeuArgIleLeuSerGluAs | 1055 |
| QY | 1696 | TTTGCTTTTAGTCGCAACGACCACTTAGAGAACCTTATCACATA----- | 1744 |
| Db | 1055 | p-----IleGluArgSerMetLeuCysGluAspLeuAlaHisAlaThrGluGlnLe | 1072 |
| QY | 1745 | -----TTCTTGTCGACGCCCAAAACGA | 1767 |
| Db | 1072 | uAsnMetLeuThrGluAlaSerLysLysHisSerGlyLeuLeuGlnSerAlaGlnGlu | 1092 |
| QY | 1768 | A-----CAAAACAAGCATCTCGAATTTCTTGAATCTGAAGATGGG----- | 1809 |
| Db | 1092 | uLeuThrLysLysGluAlaLeuIleGlnGluLeuGlnHisLysLeuAsnGlnLysLysG | 1112 |
| QY | 1810 | -GAAGTAGAACCAAGAACA---TATAGTAGTGCCCAACTCGCTCG----- | 1854 |
| Db | 1112 | uGluValGluGlnLysLysAsnGluTyrAsnPheLysMetArgGlnLeuGluHisValMe | 1132 |
| QY | 1855 | -----TATTCCATAGTGAATC | 1871 |
| Db | 1132 | tAspSerAlaAlaGluAspProGlnSerProLysThrProHisPheGlnThrHisLe | 1152 |
| QY | 1872 | CTGCTTACCTCTCCGTCCAAGAAATGAAGTAGATAGTAGAGATCAAAAGCATCTCGA | 1931 |
| Db | 1152 | uAlaLysLeuLeuGluThrGlnGluGlnGluIle-----GluAspGlyArgAlaSerLy | 1170 |
| QY | 1932 | ATGGCTAAGAGAAAAACCATTCACAAATTAAGAGAGTTTCTGTATGTTAAATGAAGAGA | 1991 |
| Db | 1170 | sThrSerLeuGluHisLeuValThrLysLeuAsnGluAspArgGluValLysAsnAlaG | 1190 |
| QY | 1992 | G-----AAAGACTGATGAACCTCTGGAATCTCCATGTCATGAGCATGG | 2036 |
| Db | 1190 | uIleLeuArgMetLysGlnGlnLeuArgGluMetGluAsnLeuArgLeuGluSerGlnG | 1210 |
| QY | 2037 | GTTTATTGCTGACAAT----- | 2052 |

Db 1210 nLeuIleGluLysanTrpLeuLeuGlnGlyGlnLeuAspAspIleLysArgGlnLysG1 1230
QY 2053 -----CAATGAATCACTGCTGTATGCTGCTTTGTAGAAATATATGACAGAAAAATAT 2105
Db 1230 uAsnSerAspGlnAsnHisProAspAsnGlnGlnLeuLysAsnGlnGlnGluGluSer11 1250
QY 2106 TAAGAGAAATTATGTCGAAACTTCATGCTTCAT-----CTAGTCAGCATGCAATGACTT 2159
Db 1250 eLysGluArgLeuAlaLysSerLysIleValGluGluMetLeuLysMetLysAlaAspLe 1270
QY 2160 TAATCTTATTAGCATTAATGTCAATAGATAAAGCTGTTACCAAGCTCCGT----- 2208
Db 1270 uGluGluValGlnSerAlaLeuTyrAsnLysGluMetGluCysLeuArgMetThrAspG1 1290
QY 2209 -----GAATGACGAGAAAATTAGAA-----AAGGGGAATCTGCTTCCCT 2249
Db 1290 uValGluArgThrGlnThrLeuGluSerLysAlaPheGlnGlnLysGluGlnLeuArgSe 1310
QY 2250 TGCAAAACGAAGAAATACTGAAGAACAAATGGACAGCAAAATGCATTTAGTGAATAATTA 2309
Db 1310 rLysLeuGluGluMetTyrGluGluArgGluArgThrSerGlnGluMetGluMetLeuAr 1330
QY 2310 CTCAAAAGAGAAAGTTTGGAAAACAGATAGTGTCTCAGGGGTTTCAAAACAGACAAAAA 2369
Db 1330 gLysGlnValGluCysLeuAlaGluGluAsnGlyLysLeuValGlyHisGlnAsnLeuHi 1350
QY 2370 ACAAAACTC 2379
Db 1350 sGlnLysIle 1353

RESULT 11

US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Alignment Scores: 1.73e-05 Length: 2482
Pred. No.: 152.50 Matches: 182
Score: 34.55% Conservative: 111
Percent Similarity: 21.46% Mismatches: 275
Best Local Similarity: 1.98% Indels: 280
Query Match: 1 Gaps: 44
DB: 1
US-09-874-162A-7 (1-4409) x US-08-328-254-6 (1-2482)
QY 112 CTCCTACTTCCACCTCCGAGCCTCATC-----GAGCAC-----ATC 150
Db 770 MetHisPheAlaGluLeuGlnGluLysPheLeuSerLeuGlnSerGluHisLysIleLeu 789
QY 151 GAGGACACCATCATGATCCACGGGTTTATAGAAAAACAA-----GAA 198
Db 790 HisAspGlnHisCysGlnMetSerLysMetSerGluLeuGlnThrTyrValAspSer 809
QY 199 TTACAGCAGCCCACTATGTTGCCCTGAGTTACATAAATAGATTTCATGACAGATGCTGCC 258
Db 810 LeuLysAlaGluAsnLeuValLeuSerThrAsnLeuArgAsnPheGlnGlyAspLeuVal 829
QY 259 CCGCAGAGCAGGAGTCCCTAAGAAAGAGATTACGCCGAAGCTCTCGCTGACTCTGCC 318
Db 830 LysGluMetGlnLeuGlyLeuGluGluValProSerLeuSer-----Ser 846
QY 319 AGCTCAGTGTCTCGAGGAATGTGTCCACTCCCGCCACGACGAGTGGAGGCTTACT 378
Db 847 SerCysVal-----ProAspSerSerSerLeuSerLeu--- 858
QY 379 CCCCCGCTGACCCCATCACCCTCTCTCTTCATTCGCGAGCAGCAGCTCCGACAGAG 438
Db 859 -----GlyAspSerSerPheTyrArg-AlaLeuLeuGluGlnTh 871
QY 439 CCAACACAGATCT---ATAGATTCTTCGAACTCGGAATCTCATAGCACCAA----- 487
Db 871 rGlyAspMetSerLeuLeuSerAsnLeuGluGlyAlaValSerAlaAsnGlnCysSerVa 891
QY 488 -----TATTTTGCACAGAACTCTTACTTACATGTCTCATC-----GAAACTCCAGA 534
Db 891 lAspGluValPheCysSerSerLeuGlnGluGluAsnLeuThrArgLysGluThrProSe 911
QY 535 ACAACATCAAAAGGAAACATTTAAAGTTGATGATATGT-----TA 576
Db 911 rAlaProAlaLysGlyValGluGluLeuSerLeuCysGluValTyrArgGlnSerLe 931
QY 577 TCAAAAGTAGAGAAATGAAAGGAGAGCAAGAAATCTCATAGCTTGTCTGACATTTGCAG 636
Db 931 uGluLysLeuGluGluLysMetGluSerGlnGlyLeuMetLysAsnLysGluIleGlnG1 951
QY 637 CTTACGTTTACTGGTTTCTTCCACAAAATG-----AT 669
Db 951 uLeuGluGlnLeuLeuSerSerGluArgGlnGlnLeuAspCysLeuArgLysGlnTyrLe 971
QY 670 AAGCCATCACAA-----ACTCAGAAAATGAACAAAATTCCTGTTACCTCGGAGCTCTG 723
Db 971 uSerGluAsnGluGlnTrpGlnGlnLysLeuThrSerValThrLeuGluMetGluSer-- 990
QY 724 CTTGTGAAAGTTTGGCACAACAAAAAGAAAGGATGTAGTTGTCCAAATAGGCAAGTCTCC 783
Db 991 -----LysLeuAlaAlaGluLys----- 996
QY 784 ACAGGTAAAAGCAGGTGCTTTGATTCTCTGACCTCAATCAACAAAAACCCGGAATTTTC 843
Db 997 -----LysGlnThrGluGlnLeuSe 1003
QY 844 CCGTCCCTTGCAGTTTCCAGTAATGAATTTGAACCTAGTAACAGCCATATGTTGAAGTCT 903
Db 1003 r-----LeuGluLeuGluValAlaArgLeuGlnLeu--GlnGlyLeuAspLeuSer 1020
QY 904 TACTCGTTGCTATTATAGTGACTCGTCCAGGAAGAGAGAGTTTAAATGAATGATTAAT 963

Db 21 SerProThrAspLeuMetThrValThrLysAsnGlnAsnIleLeuGlnSerIleSer 40
 QY 333 AGG-----GAATGTGTCACCTCC----- 350
 Db 41 ArgSerGluGluPheAspGlnAspGlyAspCysSerHisSerThrLeuValAsnGluGlu 60
 QY 351-----CCACG-----CCACAGCAGTGGAGCTTACTCC----- 380
 Db 61 GluAspProSerGlyGlyArgGlnAspTrpGlnProArgThrGluGlyValGluIleThr 80
 QY 381-----CCCCGTGACCCACCATCACCCCTCTCTTCATTCGCGACGACACTCC 431
 Db 81 ValThrPheProArgAspValSerPro-----ProGlnGluMetSer 94
 QY 432 GACAGAGCCACACAGATCTATAGATTCTTCGAACTCGGNACTCATAGCACCAATATT 491
 Db 95 Gln-GluAsp-----LeuLysGluLysAsnLeuIleAsnSerSerLe 108
 QY 492 TTGTCACAGAACTCTTACTTAC---ATGCTCTCATCGAAAC-----TCCAGAACAAACAT 542
 Db 108 uGlnGluTrpAlaGlnAlaHisAlaValSerHisProAsnGluIleGluThrValGluLe 128
 QY 543 CAAAGGAAACATTTAAAGTTGATGATATGTTATCAAAAGTACAGAAATGAAGAGAGA 602
 Db 128 uArgLysLysLysLeuThrMetArgProLeuValLeuGlnLysGluGluSerSerArgG1 148
 QY 603 GCAAGAACTCATAGCTTGTC-----GCTCATTTGCGAGCTTACGTTTAC 647
 Db 148 uLeuCysAsnValAsnLeuGlyPheLeuLeuProArgSerCysLeuGluLeuAsnIleSe 168
 QY 648 TGGTTTCTTCCACAAAATGATAGCCATCCAAAATCGAGAAAATGAACAAAATTCGT 707
 Db 168 rLysSerValThrArgGluAspAlaProHisPheLeuLysGluGlnArgLysSerG1 188
 QY 708 TACCTGGAAGTCCTGTTGAAAGTTGCGCAAAAAGAAAGAGATGAAGTTGCTC 767
 Db 188 uGluPheSerThrSerHisMetLysTySerGlyArgSerIleLysPheLeuLeuProPr 208
 QY 768 AATAAGGCAAGTCCACACA-----GGTAAAGCAGGTGCTTGTGATCTCGACCTCAA 821
 Db 208 uLeuSerLeuLeuProThrArgSerGlyValLeuThrIleProGlnAsnHisLysPhePr 228
 QY 822 TCAACAAAACCCGAAATTTCCCGTCC-----CTGCGAGTTTC 860
 Db 228 oLysGluLysGluArgAsnIleProSerLeuThrSerPheValProLysLeuSerValSe 248
 QY 861 C-----AGTAATGAATTGAACCTAGTAAC----- 885
 Db 248 rValArgGlnSerAspGluLeuSerProSerAsnGluProProGlyAlaLeuValLysSe 268
 QY 885----- 885
 Db 268 rLeuMetAspProThrLeuArgSerSerAspGlyPheIleTrpSerArgAsnMetCysSe 288
 QY 885----- 885
 Db 288 rPheProLysThrAsnHisHisArgGlnCysLeuGluLysGluGluAsnTrpLysSerLy 308
 QY 886-----AGCATATGTTGAAGTCTTACTCGTTGCT 914
 Db 308 sGluIleGluGluCysAsnLysIleGluIleThrHisPheGluLysGlyGlnSerLeuVa 328
 QY 915 A-----TTTAGTGACTCTCCAGG-----AGAAGAGATTATGAAT 956
 Db 328 lserPheGluAsnLeuLysGluGlyAsnIleProAlaValArgGluGluAspIleAspCy 348
 QY 957 GATTAAATGGAGAAACCAATGAAATATTGATGATCAATGAAGAGCTTCAGCAGAGAGAAA 1016
 Db 348 sHisGlySerLysThrArgLysProGluGluAsnSerGlnTyrrLeuSerSerArgLy 368
 QY 1017 A-----CGAAATCGTGAGATGGGAAAGACATTTGTTGACAAAT 1058
 Db 368 sAsnGluSerSerValAlaLysAsnTyrrGluGlnAspProGluIleValCysThrIlePr 388

QY 1059 GACAGATTATTGATAAAACACAGCGCTTACAGCTTTTAGATGGGATATGAA----- 1110
 Db 388 oSerLysPheGlnGluThrGlnHisSerGluIleThrProSerGlnAspGluGluMetAr 408
 QY 1111-----GTAGCCATGCCAGAA-- 1125
 Db 408 gAsnAsnLysAlaAlaSerLysArgValSerLeuHisLysAsnGluAlaMetGluProAs 428
 QY 1126-----ATGGAAGAAATGTCCAAATAGCAAGAAAGACAAACATGGGAGACTATTTCTGA 1178
 Db 428 nAsnIleLeuGluGluCysThrValLeuLysSerLeuSerSer-----ValValPheAs 446
 QY 1179 TCGGAAGAGGCTGCTCCATTTCGAAACATTTTCTCAGGACCTACGTTGCGAGTTCACTCT 1238
 Db 446 p-----AspProIleAspLysLeuProGluGlyCysSer----- 457
 QY 1239 TCGTTGGACAGGAGACCAATGATAAA---TCTACGGCTCTATTGTCACAAACCTTTGCG 1295
 Db 458-----SerMetGluThrAsnIleLysIleSerIleAlaGluArgAlaLysProGluMe 475
 QY 1296 CACTAGMAATTCAGAGAGTCTCCATCAGGAAACAAAGCTGCTTCAGTTAAACCTACTCA 1355
 Db 475 tSerArgMetValProLeuIleHis---IleThrPheProValaspGlySerProLysG1 494
 QY 1356 AACTATTGCTGTTAAAGAAATCATTTGACTACAGATCTACAAAACAGAAAA----- 1404
 Db 494 uProValIleAlaLysProSer-----LeuGlnThrArgLysGlyThrIleHi 510
 QY 1405-----GAAAGGATATCCAAATCA 1424
 Db 510 sAsnAsnHisSerValAsnIleProValHisGlnGluAsnAspLysHisLysMetAsnSe 530
 QY 1425 AAACCGA-----CAAAATTAAGAAATATTTATCATGTTCTCTATAACAA 1469
 Db 530 rHisArgSerArgArgIleThrAsnLysCysArgSerSerHisSerGluArgLysSerAs 550
 QY 1470 CAATACAAAGCAACACTGAAGCAAGAGATGACCTCGATTGCCCTTGGTGTACTCTGAA 1529
 Db 550 nIleArgThrArgLeuSerGlnLysLysThrHisMetLysCysProLysThrSerPheG1 570
 QY 1530 CTGCGCAACCTTTAGTTTACTC-----AAGCATCTTAAACTCTGCGCATAG 1577
 Db 570 yLeuLysGlnGluHisLysValLeuIleSerLysGluLysSerSerLysAlaValHisSe 590
 QY 1578 CAGATTATCTTCAACTATGTTTATCATCCAAAAGTGCT-----AGGAT 1622
 Db 590 rAsn-----LeuHisAspIleGluAsnGlyAspGlyIleSerGluProAspTrpGlnI1 608
 QY 1623 AGATGTTTATCATGATGATGTTATGATGGCTCCTATGCAGAAATCTCAGGATATTCA 1682
 Db 608 eLysSerSerGlyAsnGluPheLeuSerSerLysAspGluIleHisProMetAsnLeuAl 628
 QY 1683 TCGCAACCTGGAATTTGCTTTTAGTCGCAACGACGAGTTTAAGAGAACACCTATCACACA 1742
 Db 628 aGlnThrProGluGlnSerMetLysGlnAsn-----GluPheProProValSerAs 645
 QY 1743 TATTCTTGTGTCGAGCCCAACAGCAACAAACAGCATGCTGAA----- 1788
 Db 645 pLeuSerIleValGlu-----GluValSerMetGluGluSerThrGlyAspAr 661
 QY 1788----- 1788
 Db 661 gAspIleSerAsnAsnGlnIleLeuThrThrSerLeuArgAspLeuGlnLeuGluG1 681
 QY 1789-----TTTCTTGAATCTGAAGATGGG----- 1809
 Db 681 uLeuHisHisGlnIleProPheIleProSerGluAspSerTrpAlaValProSerGluLy 701
 QY 1810-----GAAGTAGAACAGCAAGAAACATATAGTAGTGGCCACAACTCGTCTGTA 1856
 Db 701 sAsnSerAsnLysTyrrValGlnGlnGluLysGlnAsnThrAlaSerLeuSerLysVal-- 720

```

1857 TTTCCATAGTGTACCTGCTTACCTCGTCCACAGAAATGGAGTAGTAGTGAAGA 1916
721 -----AsnAlaSerArgIleLeuThrAsnAspLeuGluPheAspSerValse 736
1917 TGAAGAGTCTGTGAATGCTTAAGAGAAAAAACCATTACAAAATTAAGAGTTTCTGA 1976
736 rAsp-----HisSerLysThrLeuThrAsnPheSerPheGlnAlaLys 750
1977 TGTTAATGAAGAGAGAGAAAGTGTGAATCTCGAATCTCCATGTCATGATGACATGG 2036
750 sGlnGluSerAlaSerSerGlnThrTyrglnTyrrp---ValHisTyrLeuAspHisAs 769
2037 GTTATTGTGTCACAAATCAATGATGCTGCTGTATGCTGTGTGTGTAGAAATTAATGACA 2096
769 pSerLeuAlaAsnLysSerIleThrTyr-----GlnMetPheGlyLys 783
2097 GAAATAATTAAGAGAAATTTATGCGAAATCTTCATGCTTCATCTAGTCAGCATCATGA 2156
783 sThrLeuSerGlyThrAsnSerIleSerGlnGluIleMetAspSerValAsnAsnGluGl 803
2157 CTTT-----AATCTTATTAGCATATATGTCATATAGATAAAGCTGTACCAAGCTCCGTA 2210
803 LeuThrAspGluLeuLeuGlyCysLeuAlaAlaGluLeuLeuAlaLeuAspGluLysAs 823
2211 AATGAGCAAAATTAAGAAAGGGGGAATCTGCTCCCTCGCAACACGAGAAATA----- 2265
823 pAsnAsnSerCysGlnLysMetAlaAsnGluThrAspProGluAsnLeuAsnLeuValle 843
2266 -----ACTGAGAGAAACAAATGGACAGCAAAATGATGATTTAGTGAATTA 2309
843 uArgTrpArgGlySerThrProLysGlu-MeGlyArgGluThrLysValLysIle- 862
2310 CTCAAAAGAGAAAGTTTGGAAACAGATAGTGTCTCAGGGGTTTCAAAACAGACAGCAAAA 2369
863 --Gln-ArgHisSerSerGly-----LeuArgIleTyrAspArgGluGluLys 877
2370 -----ACAAAACCTCTGAAAAGCTCTTAACCCC 2396
878 PheLeuIleSerAsnGluLysIlePheSerGluAsnSerLeuLysSerGluGluPro 897
2397 ATGTTATGACAAACACTCAATTTAGGGAATTCATCTCTCAAGAAATATGTTT 2456
898 IleLeuTrpThrLysGlyGluIleLeuGlyLysGlyAlaTyrGlyThrValTyrCysGly 917
2457 TTGTTT-----TTAATCATATGTTCCAAACAGGCACTGTAGATGAAGTAAT 2504
918 LeuThrSerGlnGlyGlnLeuIleAlaValLysGlnValAlaLeuAspThrSerAsn 936

```

RESULT 13

```

US-08-118-101A-2
; Sequence 2, Application US/08118101A
; Patent No. 5620892
; GENERAL INFORMATION:
; APPLICANT: Kurtz, Stephen E.
; APPLICANT: Knickerbocker, Aron M.
; APPLICANT: McCullough, John R.
; TITLE OF INVENTION: A STRAIN OF SACHAROMYCES CEREVISIAE
; TITLE OF INVENTION: EXPRESSING THE GENE ENCODING POTASSIUM TRANSPORTER MINK
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,101A

```

FILLING DATE:

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; REFERENCE/DOCKET NUMBER: DC27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-5901
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-118-101A-2

```

Alignment Scores:

```

Pred. No.: 4,66e-05 Length: 1235
Score: 146.50 Matches: 154
Percent Similarity: 31.85% Conservative: 90
Best Local Similarity: 20.10% Mismatches: 227
Query Match: 1.91% Indels: 295
DB: Gaps: 39

```

US-09-874-162A-7 (1-4409) x US-08-118-101A-2 (1-1235)

```

QY 451 TATAGATTTCTCGAACTCGGAATCTCATAGCAACCAATATTTTGCACAGAACTCTTACT 510
Db 148 PheLysMetArgThrLysThrIleLeuGluArgGluLeuThrAlaArgThrMetThr 167
QY 511 TACATGTCTCATCGAAACTCCAGAACCAACATCAAAAGGAAAAACATTTAAAGTTGTATGAT 570
Db 168 LysAsnArgThrGlyThrGlnArgThrSerTyrProArgLysGlnAlaLysThrAspAsp 187
QY 571 ATGTTATCAAAAGTA-----GAGAAAATGAAA 597
Db 188 PheGlnGluLysLeuPheSerGlyGluMetValAsnArgAspGluGlnAspSerValHis 207
QY 598 GGAGAGCAAAATCTCATAGCTTGTGTCAGCTCAATTCGAGCTTACGTTTACTGCTTTC-- 654
Db 208 SerAspGlnAsnSerHisAspIleSerArgAspSerSerAsnAsnAsnThrAsnHisAsn 227
QY 655 -----TTCCACAAAAATGATAAGCCATCAACCAAACTCAGAA 690
Db 228 GlySerSerGlySerLeuAspAspPheValLysGluAspGluThrAspAspAsnGlyGlu 247
QY 691 AATGAACAAAAATCTGTACCTCGGAAGCTCTGTTGAAAGTTGCCACACAAAAAGA 750
Db 248 TyrGlnGluAsnAsn----- 252
QY 751 AAGGATGTAAGTTGTCCAATAAGGCAAGTTCACAGATAAAAGCGGTGCCTTTGATT 810
Db 252 ----- 252
QY 811 CTTGACCTCAATCAAAACAAACCCGGAATTTCCCGTCCCTTCAGTTTCCAGTATGAA 870
Db 253 -----SerTyrSerThrValGlySerSerSerAsnThr 263
QY 871 TTT---GAACCTAGTAACGCCATATGTTGAAAGCTTACTCGTTGCTATT----- 918
Db 264 ValAlaAspGluSerLeuAsnGlnLysProLysProSerSerLeuArgPheAspGluPro 283
QY 919 ---AGATGACTCGTCCAGAGACAGAGAGTTTAAATGGAATGATTAAATGGAGAAACCAAT 975
Db 284 HisSerLysGlnArgProAlaArg----- 291
QY 976 GAAATATTGATGTCATCAATGAAGAGCTTCCAGCCAGAGAAACGAAATCGTGGAGATGG 1035
Db 292 -----ValProSerGluLysPheAlaLysArgGlySerArgAsp----- 305
QY 1036 GAAAAGACATTTGTTGTCACAAATGATGATTATGATAAAAAACAGCGCTTACAGCTTTTA 1095
Db ----- 1095

```

Db 306 -----IleSerProAlaAspMetTyrArgSerIleMetMetLeu 318
 QY 1096 GATGGGAATGAGTAGCATCGAAGAAATGGAAGATGTCATA----- 1143
 Db 319 GlnGlyLysHisGlu---AlaThrAlaGluAspGluGlyProLeuValIleGlySer 337
 QY 1144 -----AGCAAGAAAGAGCAACATGGGAG 1167
 Db 338 ProAlaAspGlyThrArgTyrLysSerAsnValAsnLysLeuLysAlaThrGly--- 356
 QY 1168 ACTATCTGTGAGGAAGAGCTCCCTCCATTCGAAACATTTCTCAGGACCTACGTTG 1227
 Db 357 -----IleAsnGlyAsnLysIleArgSerAsnGlyAsnGluSerAsnThrAsp 374
 QY 1228 CAGTTCACCTCTGTTGGACAGGAGACCAATCAATCAATCAGGCTCCTATTGCCAAA 1287
 Db 375 GlnAsnSerVal-----SerSerGlnAlaAsn-----SerThrAlaSerValSer--- 389
 QY 1288 CCTCTTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAA-----AACAAAGCCCTGTTCA 1341
 Db 390 -----AspGluSerSerLeuHisThrAsnPheGlyAsnLysValProSer 404
 QY 1342 GTTAAACCTACTCAAACT-----ATTGCTGTTAAAGAAATCATTTG 1380
 Db 405 LeuArg-----ThrAsnThrHisArgSerAsnSerGlyProIleAlaIleThrAspAsnAla 423
 QY 1381 ACTACAGATCTACAAACAAGAAAGAAAGGATACTCAATGAAACCCGACAAATA 1440
 Db 424 GluThrAsp-----LysLysHisGlyProSer----- 432
 QY 1441 AGAATATTTATGTTTCTCTATAACAACAATCAAGG-----CAACAAACTGAA 1491
 Db 433 -----IleGlnPheAspIleThrLysProArgLysIleSerLysArgValSer 449
 QY 1492 GCAAGAGATGACCTGCATGCTGCTGTTGTTACTCTGAACTCGCGCAACTTTATAGT--- 1548
 Db 450 ThrPheAspAspLeuAsn---ProLysSerValLeuTyrArgLysLysAlaSerLys 468
 QY 1549 -----TTACTCAAGCATCTTAAACTCTGCCATACAGAGATTTATCTCAACTATGTTAT 1602
 Db 469 LysTyrLeuMetLysHisPhe----- 475
 QY 1603 CATCCAAAGTGCTAGATAGATGTTTCTATC---RATGAGTGTATGATGCTCTCAT 1659
 Db 476 ---ProLysAlaArgArgIleArgGlnGlnIleLysArgArgLeuSerThrGlySerIle 494
 QY 1660 GCAGGAATCCTCAGGATATTCTATCGCCCACTGGATTGCTTTAGTCGCAACGGACCA 1719
 Db 495 GluLysAsn-----SerSerAsnAsnVal 502
 QY 1720 GTTAAAGAACACCTATCACATATTTCTTGTGTGCGCGCCAAACAAAGCAAGC 1779
 Db 503 SerAspArgLysProIleThr----- 510
 QY 1780 ATGCTCTGAA-----TTTCTT 1794
 Db 511 MetAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 530
 QY 1795 GAATCTGAAGATGGGAAGTAGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGC 1854
 Db 531 AspAsnGluSerGlyAsp---GluAspGluArgValGlnGlnSerGluPro----- 546
 QY 1855 TATTTCCATGATGATACCTGTTTACCTCTCCGTCACAGAAATGGA----- 1902
 Db 547 -----HisSerAspSerGluLeuLysSerHisGlnGlnGlnGlnLysHisGlnLeu 564
 QY 1903 -----GTAGATAGTGAAGATGAAAGGAT 1926
 Db 565 GlnGlnAsnLeuHisArgMetTyrLysThrLysSerPheAspAspAsnArgSerArgAla 584
 QY 1927 CTGATCGCTAAGAGAAACCAATTTACAAATTTAGAGTTTCTGATGTTAATGAA 1986
 Db 585 ValProMetGluArgSerArgThrIle---AspMetAlaGluAlaLysAspLeuAsnGlu 603

QY 1987 -----GGAGAGAAAGAAAGTGTATGAAGTCTGCAATCTCCAT----- 2022
 Db 604 LeuAlaArgThrProAspPheGlnLysMetValTyrGlnAsnThrLysAlaHisArg 623
 QY 2023 -----GTCATGAAGCATGGG---TTT 2040
 Db 624 LysLysProAsnPheArgLysArgGlyTrpAsnAsnLysIlePheGluHisGlyProTyr 643
 QY 2041 ATTGCTGCAATCAATCAATCAATCCCTGTATGCTGTTGTAGAAAATTATGGACAGAA 2100
 Db 644 AlaSerAspSerAspArgAsnTyrProAspAsn-----SerAsnThrGly----- 658
 QY 2101 ATAATTAAGAAGATTTATGTCGAAACCTTCATGCTTCTATCTAGTCAGCATGCATGACTTT 2160
 Db 659 -----AsnSerIleLeuHisTyrAlaGlu----- 666
 QY 2161 AATCTTATGATATGTCATATAGATAAAGCTGTACCAAGCTCCGTGAATGCAGCAA 2220
 Db 667 -----SerIleLeuHisAspGlySer-----His 675
 QY 2221 AAATTTAGAAAAGGGGCAATCTCTCCCTCGCAACGAAAGAAATAACTGAAGAACAAAAT 2280
 Db 676 LysAsnGlySerGluGluAlaSerSerAspSerAsnGluAsnIleTyrSerThrAsnGly 695
 QY 2281 GGGACA-----GCAAAATGGATTAGTGAATAAATACTCAAAAGAGAAAGCT--- 2325
 Db 696 GlySerAspHisAsnGlyLeuAsnAsnTyrProThrTyrAsnAspAspGluGluGlyTyr 715
 QY 2326 -----TTGGAACACAGATAGTCTCAGGGGTTTCAAAACAG 2361
 Db 716 TyrGlyLeuHisPheAspThrAspTyrAspLeuAspProArgHisAspLeuSerLysGly 735
 QY 2362 AGCAAAAAACAAAAACTC 2379
 Db 736 SerGlyLysThrTyrLeu 741
 RESULT 14
 US-09-418-710-21
 ; Sequence 21, Application US/09418710
 ; Patent No. 6596482
 ; GENERAL INFORMATION:
 ; APPLICANT: Jones, Michael H.
 ; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
 ; FILE REFERENCE: 06501-042001
 ; CURRENT APPLICATION NUMBER: US/09/418,710
 ; PRIOR FILING DATE: 1999-10-15
 ; PRIOR APPLICATION NUMBER: PCT/JP98/01783
 ; PRIOR FILING DATE: 1998-04-17
 ; PRIOR APPLICATION NUMBER: JP 9/310027
 ; PRIOR FILING DATE: 1997-10-24
 ; PRIOR APPLICATION NUMBER: JP 9/116570
 ; PRIOR FILING DATE: 1997-04-18
 ; NUMBER OF SEQ ID NOS: 73
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 21
 ; LENGTH: 1972
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-418-710-21
 Alignment Scores:
 Pred. No.: 6,548-05 Length: 1972
 Score: 146.00 Matches: 154
 Percent Similarity: 32.90% Conservative: 126
 Best Local Similarity: 18.10% Mismatches: 314
 Query Match: 1.90% Indels: 257
 DB: 4 Gaps: 33
 US-09-874-162a-7 (1-4409) x US-09-418-710-21 (1-1972)
 QY 139 ATCGAGCATCGAGAGCAACCATCGATCCAGATCCACGGGTTTATAGAAAAACAGAA 198

QY 1687 CAACCTGGATTTGCTTTTGTAGTCGCAACGGACCGAGTAAAGAACACCTATCATCATATT 1746
Db 398 -----ArgThrAspLeuProAsnSer 404
QY 1747 CTGTGTGTCAGGCGCAAAACGACAAAGCAAGCATGCTGTAATTTCTTGAATCTCAAGAT 1806
Db 405 ValVal-----AsnAspAsp 409
QY 1807 GGGGAAGTAGAACAGCAA-----AGAACAATATAGTAGTGGCCACAATCGTCTGTAT 1857
Db 410 IleGluTyrAsnGlnLysSerAspAsnAsnSerTyrSerThr-----AsnAsnLeuTyr 427
QY 1858 TTCCATAGTGATACCTGCTTACTCTCCGTCACAGAANAATGGAAGTAGTAGAAGAT 1917
Db 427 ----- 427
QY 1918 GAAAGGATCTCTGAATGGCTAAGAGAAAAAACCATTACAAAATGGAAGATTTCTGTAT 1977
Db 428 -----AsnAsnIleAsnMetThrGlnAsnGlnAspAsnAsnAsn 440
QY 1978 GTTAATGAAGGAGAGAAAGAGTGAATAACTCTGGAATCTCCATGTCATGAAGCATGG 2037
Db 441 ValAsn-----IlePheLys-----HisValGlnAsnAspSer 451
QY 2038 TTTATTGCTGACAATCAAAATGAATCATGCTGATGCTGTTGTAGAA----- 2085
Db 452 PheGlnCysPheAsnSerAsnAsn-----LeuTyrIleGluLysAspIleLys 467
QY 2086 -----AATATGGACAGAAAAATAATTAAGAAATTTATGT-----CGAACTTCATG 2133
Db 468 GluAsnAsnIleSerGln-----IleAsnArgLysLeuLysThrLysArgAsnPhe--- 484
QY 2134 CTTTCATCTAGTCAGATGATGACTTTAATCTATATAGCATTAATCTCAATAGATAAAGCT 2193
Db 485 -----ThrLysLysSerArgLysLysIleAsnThrValThrTyrLeuGlnIleAspLys--- 501
QY 2194 GTTACCAAGCTCGTGAATGAGCAAAATAATTAAGAAAGGGGGAATCTGCTTCCCTGCA 2253
Db 502 ValIleLysIle-LeuLysCysLys----- 509
QY 2254 AACGAGAAATCACTGAGAAC-----AAATGGACAGCAAAATGATTTAGTGAAT 2307
Db 510 -----LysLysTyrIleLysHisIleLysLysMetLysTyrMetAsnAsnPheGlnAsnPh 528
QY 2308 AACTCAAAAGAGAAAGCTTTGAAACAGATAGTGTCTCAGGGTTTCAAAACAGAGCA-- 2365
Db 528 elysLysLeuLysLysLeu-----GlnLysPheGlnAsnAlaSerPh 542
QY 2366 -----AAAAACAATACTCTGAAAAGCTCTAACCCCATGTTATGGA 2406
Db 542 eglLeuLysIleAsnLysIleAsnLysAsnIleArgArgLeu-----AsnLysLeuLys 560
QY 2407 CAACACTGAATTTACATTTTAGGGAATTCATCC-----TCTAA 2445
Db 560 sLysArgLysAsnHisSerIleAsnIleThrProValThrSerIleGluTrpLeuAsnAs 580
QY 2446 GAATTTATGTTTTGTTTTT---AATCATATGTTCCAAACAGCAGCACTGTAGTGAAGTAA 2502
Db 580 nSerTyrThrPheAspPheIleAsnAsnSerIleGlnSer----- 593
QY 2503 ATGATTTCAACAAGATATTGTCAGGGT-----TCTACTTCATTCATTATGC 2553
Db 594 -----ThrSerTyrProTrpLysAsnLysCysAspAlaThrIleArgAsnHisLe 610
QY 2554 AGCATTACATGATATCACTTTTATTGATGTCATTAAACATCTCTGTTACTTTAAGCATGA 2613
Db 610 uHisLeuHis-----AsnValIleIleAspLysAsnAs 621
QY 2614 AAGCAATATTTCAAGATATTTTAAACTCAACAATGTCATCAA----- 2659
Db 621 nLysThrTyrPheMetLysAsnLeuValGluAsnArgIleValArgAsnIleIleSerLy 641
QY 2660 -----TATGTTGAATTGATCTAGAAATATTTCATATATAAT 2697

Db 641 sGlnLysLysCysGlnSerLeuTyr-LysAsnLysGlnAsnValTyrPheCysTyrLysA 661
QY 2698 CAGAAATTTTTCATTTATGACCGGCTGTTTTTCTACTTTGTAATTTGTGAGACATTTTC 2757
Db 661 snAsnPheSerLeuLeuLysSerSerIle----- 670
QY 2758 TTGGGAGGGGAAAATTTGGAATGTTCCCTTTTGTAGAAATTTGAAGTGGTCTTCATATGTC 2817
Db 671 -----LeuLysPheIleCysC 676
QY 2818 AACTACAGAAAGGAAAAAATA-----GAAATTTGAAGATTTTATGAAATTTATATTG 2871
Db 676 ysIleLysThrLeuLysLysMetPheAsnAlaPheThrAsnSerThrTyrAsnThrLysP 696
QY 2872 CATTACTATTGTCAGTCAAACTTTTGATGCTTTTGTGAAATCATTTGTCAATTCGGAAT 2931
Db 696 helIeLeuPheLeuIleSer-----TyrMet-AsnLysMetLeuIleLysAsn 711
QY 2932 GAAAAATTAATGTAATTTTACATTAATGATGCTTCTTTTCAAA----- 2976
Db 712 LysLysLeu-----LysPhe--ValLysLeuPheLeuIleGlnThrAlaIleGluAla 728
QY 2977 -----TTAAAAATAGCACTTCTTCATCTTATGCTGCTTTTGAAGATAT 3021
Db 729 PheArgTyrAlaArgIlePheAsnGlnAspSerPheTyrProCysLeuGlnHisPhe 748
QY 3022 ---TAAATTTTCATTTGTCAGAGAAATGCTATGTTGCTTTTATAAGATTACAGACCA 3078
Db 749 ArgLysIleLysLysArgLeuIleAsnLysTyrLysIleGly----- 762
QY 3079 TTTGTTTTCATGTCGATAATTTTAGTCATGCTCACCCCGTATGTTTTTTTTTTTAAAC 3138
Db 763 -----HisAsn 764
QY 3139 TTGAACATTTTGTGCTTTTGTGTTTTTCTTTTAAAT 3174
Db 765 LysAsnLeuLeuArgGluPhePheLeuPheAsn 776

Search completed: August 25, 2004, 19:07:37
Job time : 205 secs

Length: 776

```

; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-162A-8

```

| | | | |
|------------------------|---------|---------------|-----|
| Score: | 4078.00 | Matches: | 776 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 53.03% | Indels: | 0 |
| DB: | 9 | Gaps: | 0 |

| | | | |
|--|------|--|------|
| US-09-874-162A-7 (1-4409) x US-09-874-162A-8 (1-776) | | | |
| QY | 52 | ATGACAGGATCGCCGCGCTCTCTCTCAATACCTGCCGATTCGGGGTGGCGA | 111 |
| DB | 1 | MetThrGlyLeuAlaAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGly | 20 |
| QY | 112 | CTCCATCTCCCGACCTCGCGACCTCATCGAGACATCGAGACCAACCATCGATACA | 171 |
| DB | 21 | LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThr | 40 |
| QY | 172 | GATCCACGGTGTAGAAAAACAAGATTACAGAGCAACCTATGTTGCCCTGAGTTAC | 231 |
| DB | 41 | AspProArgValLeuGluGlyGlnGluGlnGlnProThrTyrValAlaLeuSerTyr | 60 |
| QY | 232 | ATAAATAGATTATGACAGATGCTCCCGCGAGAGCAGAGTCCCTAAAGAAAGATT | 291 |
| DB | 61 | IleAsnArgPheMetThrAspAlaAlaArgGluGlnGluSerLeuLysLysLysIle | 80 |
| QY | 292 | CAGCGAGAGCTCGCTGACTCTGTCAGCTCAGTGTCTCGAGGAATGTCTCCACTCC | 351 |
| DB | 81 | GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro | 100 |
| QY | 352 | CCACGCCACAGAGTGGAGGCTTACTCCCGCGTGACCCACCCATCACCCCTCTCT | 411 |
| DB | 101 | ProArgHisSerSerGlySerLeuThrProValThrProProLleThrProSerSer | 120 |
| QY | 412 | TCATTCCGAGAGCACTCCGACAGAGCCAAACAGATCTATAGATTCTTCGAACTCGG | 471 |
| DB | 121 | SerPheArgSerSerThrProThrGluProThrGlnIleTyrArgPheLeuArgThrArg | 140 |
| QY | 472 | AATCTCATAGACCAATATTTTGCACAGAACTCTTACTTACATGTCTCATCGAAACTCC | 531 |
| DB | 141 | AsnLeuIleAlaProIlePheLeuHisArgThrLeuThrTyrMetSerHisArgAsnSer | 160 |
| QY | 532 | AGAAACAACATCAAAAGAAACATTTAAAGTTGATGATATGTTATCAAAAGTAGAGAA | 591 |
| DB | 161 | ArgThrAsnIleLysArgLysThrPheLysValAspAspMetLeuSerLysValGluLys | 180 |
| QY | 592 | ATGAAGCAGAGCAAGATCTCATAGCTTGCAGCTCATTTGCGAGCTTACGTTTACCTGGT | 651 |
| DB | 181 | MetLysGlyGluGlnGlnSerHisSerLeuSerAlaHisLeuGlnLeuThrPheThrGly | 200 |
| QY | 652 | TTCTTCCCAAAATGATAGCCCATCACCAACTCAGAAAAATGAACAAAATTTCTGTTACC | 711 |
| DB | 201 | PhePheHisLysAsnAspLysProSerProAsnSerGluAsnGluGlnAsnSerValThr | 220 |
| QY | 712 | CTGGAGTCTCTGTTGAAAGTTGCCACAAAAAGAGAGATGTAAGTTGTCGAATA | 771 |
| DB | 221 | LeuGluValLeuLeuValLysValCysHisLysLysArgLysAspValSerCysProIle | 240 |
| QY | 772 | AGGCAAGTTCCACAGGTAAAGAGCAGGTGCTTTGATTTCTGACCTCAATCAAAACAAA | 831 |
| DB | 241 | ArgGlnValProThrGlyLysLysGlnValProLeuIleProAspLeuAsnGlnThrLys | 260 |
| QY | 832 | CCCGAAATTTCCGCTCCCTTCCAGTTTCCAGTAAATGAACTTGAACCTAGTAACAGCAT | 891 |
| DB | 261 | ProGlyAsnPheProSerLeuAlaValSerSerAsnGluPheGluProSerAsnSerHis | 280 |
| QY | 892 | ATGGTGAAGTCTTACTCTGCTATTATAGAGTGACTCGTCCAGGAAGAAGAGAGTTTAAAT | 951 |
| DB | 281 | MetValLysSerTyrSerLeuLeuPheArgValThrArgProGlyArgArgGluPheAsn | 300 |
| QY | 952 | GGAATGATTAATGGAGAAACCAATGAAATTAATGATGTAATGAAGAGCTTCCAGCCAGA | 1011 |
| DB | 301 | GlyMetIleAsnGlyGluThrAsnGluAsnIleAspValAsnGluGluLeuProAlaArg | 320 |
| QY | 1012 | AGAAAACGAAATCGTGAGGATGGGAAAAAGACATTTGTGTGCACAAATGACAGATTGAT | 1071 |

| | | | |
|----|------|---|------|
| DB | 321 | ArgLysArgAsnArgGluAspGlyGluLysThrPheValAlaGlnMetThrValPheAsp | 340 |
| QY | 1072 | AAAAACAGCGGCTTACAGCTTTTAGATGGGAATATGAAGTAGCATCGAGAAATGAAA | 1131 |
| DB | 341 | LysAsnArgArgLeuGlnLeuLeuAspGlyGluTyrGluValAlaMetGlnGluMetGlu | 360 |
| QY | 1132 | GAATGTCCCAATAGCAAGAAAAAGACACATCGGAGACTATTCTTGATGGGAAGGCTG | 1191 |
| DB | 361 | GluCysProIleSerLysLysArgAlaThrTrpGluThrIleLeuAspGlyLysArgLeu | 380 |
| QY | 1192 | CCTCCATTCGAAACATTTTCTCAGGACCTAGCTTGAGTTCAGTCTCTCTGTTGGACGA | 1251 |
| DB | 381 | ProProPheGluThrPheSerGlnGlyProThrLeuGlnPheThrLeuArgTrpThrGly | 400 |
| QY | 1252 | GAGACCAATGATAAATCTACGGCTCTTATTTGCCAAACCTCTTGCACACTAGAAATTCAGAG | 1311 |
| DB | 401 | GluThrAsnAspLysSerThrAlaProIleAlaLysProLeuAlaThrArgAsnSerGlu | 420 |
| QY | 1312 | AGTCTCCATCAGAAAAAACAAGCTGGTTCAGTTAAACCTACTCAAACTATTCTCTTTAAA | 1371 |
| DB | 421 | SerLeuHisGlnGluAsnLysProGlySerValLysProThrGlnThrIleAlaValLys | 440 |
| QY | 1372 | GAATCAATTCAGTACAGATCTACAAACAGAAAAAGAAAGGATACCTCCAAATGAAACCGA | 1431 |
| DB | 441 | GluSerLeuThrThrAspLeuGlnThrArgLysGluLysAspThrProAsnGluAsnArg | 460 |
| QY | 1432 | CAAAATTAAGAAATATTTATCAGTTTCTTATAACAACAATACAAAGCAACAACACTGAA | 1491 |
| DB | 461 | GlnLysLeuArgIlePheTyrGlnPheLeuTyrAsnAsnAsnThrArgGlnGlnThrGlu | 480 |
| QY | 1492 | CAAGAGATGACCTGCATTCCTGCTGCTACTCTGAACTCGCACTCGCAAACTTTATAGTTTA | 1551 |
| DB | 481 | AlaArgAspAspLeuHisCysProTyrCysThrLeuAsnCysArgLysLeuTyrSerLeu | 500 |
| QY | 1552 | CTCAAGCATCTTAAACTCTGCCATAGCAGATTTATCTTCAACTATGTTTATCATCAAAA | 1611 |
| DB | 501 | LeuLysHisLeuLysLeuCysHisSerArgPheIlePheAsnTyrValTyrHisProLys | 520 |
| QY | 1612 | GGTCTAGATAGATGTTTCTATCAATGAGTGTATGATGCTCTCTATGCGAGGAATCCT | 1671 |
| DB | 521 | GlyAlaArgIleAspValSerIleAsnGluCysTyrAspGlySerTyrAlaGlyAsnPro | 540 |
| QY | 1672 | CAGGATATTCATCGCAACCTCGATTGCTTTTCTGCTCGCAACGAGACCTAGTAAAGAGACA | 1731 |
| DB | 541 | GlnAspIleHisArgGlnProGlyPheAlaPheSerArgAsnGlyProValLysArgThr | 560 |
| QY | 1732 | CTATCACACATATTCCTGTGTGCGAGCCAAACGAAACGAAAGCAAGCATGTCTGAATTT | 1791 |
| DB | 561 | ProIleThrHisIleLeuValCysArgProLysArgThrLysAlaSerMetSerGluPhe | 580 |
| QY | 1792 | CTTGAATCTGAAGTGGGAAGTAGACAGCAAGAAACATATAGTAGTGGCCACAACTCGT | 1851 |
| DB | 581 | LeuGluSerGluAspGlyGluValGlnGlnGlnArgThrTyrSerSerGlyHisAsnArg | 600 |
| QY | 1852 | CTGTATTTCCATPAGTATACCTGCTTACTCTCTCTCTCCGTCACAGAAATGGAAGTAGATAGT | 1911 |
| DB | 601 | LeuTyrPheHisSerAspThrCysLeuProLeuArgProGlnGluMetGluValAspSer | 620 |
| QY | 1912 | GAAGATGAAGAAGATCCTGAAATGGCTAGAGAAAAAACCATTACAAATTCAGAGATTT | 1971 |
| DB | 621 | GluAspGluLysAspProGluThrLeuArgGluLysThrIleThrGlnIleGluGluPhe | 640 |
| QY | 1972 | TCTGATGTTAATGAAGAGAGAAAGATGATGAACCTCGAATCTCCATGTCATGTAAG | 2031 |
| DB | 641 | SerAspValAsnGluGlyGluLysGluValMetLysLeuTyrAsnLeuHisValMetLys | 660 |
| QY | 2032 | CATGGTTTATTCCTGACAAATCAAAATGAATCATGCTGTATGCTGTTTGTAGAAAAATAT | 2091 |
| DB | 661 | HisGlyPheIleAlaAspAsnGlnMetAsnHisAlaCysMetLeuPheValGluAsnTyr | 680 |
| QY | 2092 | GCACAGAAAAATTAATGAAGAAATTTATGTCGAAACTTCATGCTTCATCTAGTCAGCATG | 2151 |

1876 TTACCTCTCCGTCACAAAGAAATGGAAGTAGATAGTGAAGATGAAAGGATCCTGAATGG 1935
 572 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAspProGluTrp 591
 1936 CTAAGAGAAAAACCATTAACAATGGAAGTCTTCTGATGTTAATGAAGAGAGAGAAA 1995
 592 LeuArgGluLysThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluLys 611
 1996 GAAGTGCATGAAACTCTGGAATCTCCATGTCATGAAGCATGGTTTATTGCTGACAATCAA 2055
 612 GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 631
 2056 ATGAATCATGCTGTATGCTGTTCTAGAAAATTATGACAGAGAAAATAATTAAGAAGAAT 2115
 632 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleLeuLysAsn 651
 2116 TTATGTCGAACTTCATGCTTCATCTAGTCAGCATGCGATGCTTAATCTTATTAGCATA 2175
 652 LeuCysArgAsnPheMetLeuHisValSerMetHisAspPheAsnLeuIleSerIle 671
 2176 ATGTCATAGATAAAGCTGTTACCAAGCTCCGTGAAATGCAAGCAAAAATTAGAAAAGGG 2235
 672 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 691
 2236 GAATCTGCTCCCTGCAAGCAAGAAATTAAGTGAAGCAAAAATGGGACAGCAAAATGGA 2295
 692 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 711
 2296 TTTAGTGAATTAATCTCAAAAGAGAAAGCTTTGGAACAGATAGTGTCTCAGGGGTTTCA 2355
 712 PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 731
 2356 AAACAGAGCAAAAAACAATACTC 2379
 732 LysGlnSerLysLysGlnLysLeu 739

RESULT 3
 US-09-764-864-931
 ; Sequence 931, Application US/09764864
 ; Patent No. US20020132753A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT223
 ; CURRENT APPLICATION NUMBER: US/09/764,864
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 1792
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 931
 ; LENGTH: 388
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-764-864-931

Alignment Scores:
 Pred. No.: 1,8e-169 Length: 388
 Score: 2043.00 Matches: 386
 Percent Similarity: 99.48% Conservations: 0
 Best Local Similarity: 99.48% Mismatches: 2
 Query Match: 26.57% Indels: 0
 DB: 9 Gaps: 0

US-09-874-162A-7 (1-4409) x US-09-764-864-931 (1-388)

1216 GGACCTAGCTTCCAGTTTCACTCTGTTGGACAGAGACCAATGATAAATCTACGGCT 1275
 1 GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAla 20
 1276 CCTATTGCAAAACCTCTTGGCCCATAGAAATTCAGAGAGTCTCCATCAGGAAACCAAGCCT 1335
 21 ProIleAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro 40

1336 GGTTCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAATCATTGACTACAGATCTACAA 1395
 41 GlySerValLysProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGln 60
 1396 ACAAGAAAAAGAGATACTCCAAATGAATGAAACCCGACAAAAATTAAGAAATATTTATCAG 1455
 61 ThrArgLysGluLysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheThrGln 80
 1456 TTTCTCTATAACAACAATCAAGGCAACAACTGAAGCAAGAGATGACCTGATTCGCTCCT 1515
 81 PheLeuTyrAsnAsnAsnThrArgGlnGlnThrGluAlaArgAspAspLeuHisCysPro 100
 1516 TGTGTACTCTGAACCTGCCGCAAACTTTATAGTTTACTCAACATCTTAAACTCTGCCAT 1575
 101 TrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysLeuCysHis 120
 1576 AGCAGATTTTATCTTCAACTATGTTTATCATCCAAAAGGTGCTAGGATAGATGTTTCTATC 1635
 121 SerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIle 140
 1636 AATGAGTGTATTGATCGCTCTCTATGAGGAAATCCTCAGGATATTTCATCGCAACCTGGA 1695
 141 AsnGluGlyTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGly 160
 1696 TTTGCTTTTGTAGTCGCAACGAGACAGTAAAGAGAACACCTATCACACATATTTCTGTGTC 1755
 161 PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys 180
 1756 AGSCCAAAACGAAACAAACCAAGCATGCTGAAATTTCTGAACTGGAAGATGGGGAAGTA 1815
 181 ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal 200
 1816 GAACAGCAAAAGAACATATAGTAGTGGCCCAATCTGTTATTTCCATAGTATACCTGTC 1875
 201 GluGlnGlnArgThrTyrSerSerGlyHisAsnArgLeuTyrPheHisArgAspThrCys 220
 1876 TTACCTCTCCGTCACAAAGAAATGGAAGTAGATAGTGAAGATGAAAGGATCCTGAATGG 1935
 221 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAspProGluTrp 240
 1936 CTAAGAGAAAAAACCATTAACAATGGAAGTCTTCTGATGTTAATGAAGAGAGAGAAA 1995
 241 LeuArgGluLysThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluLys 260
 1996 GAAGTGCATGAAACTCTGGAATCTCCATGTCATGAAGCATGGTTTATTGCTGACATCAA 2055
 261 GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 280
 2056 ATGAATCATGCTGTATGCTGTTCTAGAAAATTATGACAGAGAAAATAATTAAGAAGAAT 2115
 281 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleLeuLysLysAsn 300
 2116 TTATGTCGAACTTCATGCTTCATCTAGTCAGCATGCTTAAATCTTATTAGCATA 2175
 301 LeuCysArgAsnPheMetLeuHisValSerMetHisAspPheAsnLeuIleSerIle 320
 2176 ATGTCATAGATAAAGCTGTTACCAAGCTCCGTGAAATGCAAGCAAAAATTAGAAAAGGG 2235
 321 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 340
 2236 GAATCTGCTCCCTGCAAGCAAGAAATAATTAAGAAGCAAAAATGGGACAGCAAAATGGA 2295
 341 GluSerAlaSerProAlaAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGly 360
 2296 TTTAGTGAATTAATCTCAAAAGAGAAAGCTTTGGAACAGATAGTGTCTCAGGGGTTTCA 2355
 361 PheSerGluIleAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 380
 2356 AAACAGAGCAAAAAACAATACTC 2379
 381 LysGlnSerLysLysGlnLysLeu 388

RESULT 4

```
US-09-764-864-911
; Sequence 911, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 911
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-911

Alignment Scores:
Pred. No.: 9,43e-122 Length: 289
Score: 1497.00 Matches: 283
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.65% Mismatches: 0
Query Match: 19.47% Indels: 0
DB: 9 Gaps: 0

US-09-874-162A-7 (1-4409) x US-09-764-864-911 (1-289)
QY 1528 AACTGCCGCAACTTATAGTTTCTCAAGCATCTTAACTCTGCCATGACAGATTATC 1587
Db 6 AspCysArgLeuTyrSerLeuLeuLysHisLeuLysLeuCysHisSerArgPheile 25
QY 1588 TTCACATATGTTTATCATCCAAAGGTCTAGAGTAGATGTTCTCATAGAGTGTAT 1647
Db 26 PheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIleAsnGluCysTyr 45
QY 1648 GATGCTCTCTATCGAGGAATCCTCAGGATATTCATCCCACTGGATTTCCTTTAGT 1707
Db 46 AspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGlyPheAlaPheSer 65
QY 1708 CCAACGACCACTTAAGAGAACCTATCACACATATCTCTGTGAGGCCAAAGCA 1767
Db 66 ArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCysArgProLysArg 85
QY 1768 ACAAAAGCAAGCATGCTCTGAATTTCTTGAATCTGAAGATGGGAAGTAGAACAGCAAGA 1827
Db 86 ThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluValGlnGlnArg 105
QY 1828 ACATATAGTAGTGGCCACAATCGTCTGTATTTCCATAGTAGTACTGTCTTACCTCTCGT 1887
Db 106 ThrTyrSerSerGlyHisAsnArgLeuTyrPheHisSerAspThrCysLeuProLeuArg 125
QY 1888 CCACAAAGAAATGGAAGTAGATAGTCAAGATGAAGATGAAGATCTGTATGCTTAAGAGAAA 1947
Db 126 ProGlnGluMetGluValAspSerGluAspGluLysAspProGluTrpLeuArgGluLys 145
QY 1948 ACCATTACAAATTAAGAGATTTCTGTATGTTAATGAAGAGAGAAAGAAAGATGATGAAA 2007
Db 146 ThrIleThrGlnIleGluPheSerAspValAsnGluGlyGluLysGluValMetLys 165
QY 2008 CTTCTGAATCTTCATGTCATGAGCATGGGTTTATGCTGACATCAATCAATGATCATGCC 2067
Db 166 LeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGlnMetAsnHisAla 185
QY 2068 TGTATGCTGTTTGTAGAAAATATGACAGAGAAAATATTAAGAGAAATTTATGTCGAAAC 2127
Db 186 CysMetLeuPheValGluAsnTyrGlyGlnLysIleIleLysLysAsnLeuCysArgAsn 205
QY 2128 TTCATGCTTCATCTAGTCAGCATGATGATCTTTAATCTTATAGCATTAATGTCATAGAT 2187
Db 206 PheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIleMetSerIleAsp 225
QY 2188 AAAGCTGTTTACCAGCTCGTGAATGACGACAAAATTTAGAAAAGGGGGATCTGCTTCC 2247
```

```
Db 226 LysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGlyGluSerAlaSer 245
QY 2248 CCTGCAACGAAGAAATACTCTGAGACACAAATGGGACACAAATGGATTTAGTGAATT 2307
Db 246 ProIleAsnGluGluIleThrGluGluGlnAsnGlyThrAlaAsnGlyPheSerGluIle 265
QY 2308 AACTCAAAAGAGAAAGCTTTTGGAAACAGATAGTGTCTCAGGGGTTTCAAAACAGACAAA 2367
Db 266 AsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSerLysGlnSerLys 285
QY 2368 AACACAAAACCTC 2379
Db 286 LysGlnLysLeu 289

RESULT 5
US-09-764-864-1369
; Sequence 1369, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1369
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (9)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (16)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (31)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (46)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (121)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (283)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1369

Alignment Scores:
Pred. No.: 6,05e-118 Length: 292
Score: 1453.50 Matches: 280
Percent Similarity: 96.25% Conservative: 2
Best Local Similarity: 95.56% Mismatches: 9
Query Match: 18.90% Indels: 2
DB: 9 Gaps: 1

US-09-874-162A-7 (1-4409) x US-09-764-864-1369 (1-292)
QY 748 AGAAGATGATTAAGTTGTCCCAATAGGCAAGTCCACAGTAAAGCAGGTGCTTTG 807
Db 1 LysLysAspValSerCysProIle***GlnValProThrGlyLys***Gln***ProLeu 20
QY 808 ATTCTCGACCTCAATCAACAAACCCGGAATTTCCCGTCCCTTCAGTTTCCAGTAAT 867
Db 21 AsnProAspLeuAsnGlnThrLysProGly***PheProSerLeuAlaValSerSerAsn 40
```

```
QY 868 GAATTGAACTAGTAACAGCCATATGCTGAAGTCTTACTCGTTGCTATTATAGAGTCACT 927
Db 41 GluPheGluProSer***SerHisMetValIysSerLeuLeuPheArgValThr 60
QY 928 COTCCAGGAAGAGAGATTAAATGAATGATTAATGAGAAACCAATGAAATATTGAT 987
Db 61 ArgProGlyArgGluPheAsnGlyMetIleAsnGlyGluThrAsnGluAsnIleAsp 80
QY 988 GTCAATGAGAGCTTCCAGCCAGAGAAACGAATCGTAGGATGGGAAACACATTT 1047
Db 81 ValAsnGluGluLeuProAlaArgArgLysArgAsnArgGluAspGlyGluYsthrPhe 100
QY 1048 GTTGACCAAAATGACAGTATTTGATAAAACAGGGCTTTACAGCTTTTAGATGGGAATAT 1107
Db 101 ValAlaGlnMetThrValPheAspLysAsnArgArgLeuGlnLeuLeuAspGlyGluYst 120
QY 1108 GAAGTAGCCATCCAGGAATGGAAGAAATGCAATTAACAGAAAGAAAGCAATGGGAG 1167
Db 121 ***ValAlaMetGlnGluMetGluGluCysProIleSerLysLysArgAlaThrTIPGlu 140
QY 1168 ACTATTCTTGAGGAAAGAGGCTGCTCCATTTCGAAACATTTCTCAGGGACCTACGTTG 1227
Db 141 ThrIleLeuAspGlyLysArgLeuProPheGluThrPheSerGlnGlyProThrLeu 160
QY 1228 CAGTTCACTCTTCGTTGACAGGAGAGACCAATGATAAATCTACGGCTCCTATTGCCAAA 1287
Db 161 GlnPheThrLeuArgTIPThrGlyGluThrAsnAspLysSerThrAlaProIleAlaLys 180
QY 1288 COTCTTGCACTAGAAATTCAGAGAGTCTCCATCAGGAAACCAAGCTGGTTTCAGTTAAA 1347
Db 181 ProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysProGlySerValLys 200
QY 1348 COTACTCAAACTATTGCTGTTAAAGAATCATTGACTCAGATCTACAAACAAAGAAAGAA 1407
Db 201 ProThrGlnThrIleAlaValLysGluSerLeuThrThrAspLeuGlnThrArgLysGlu 220
QY 1408 AAGGATCTCCAAATGAAACCGACAAATTAAGATATTTATCAGTTCTCTATAAC 1467
Db 221 LysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheThrGlnPheLeuYsthr 240
QY 1468 AACATACAAAGCAACAACTGAAGCAAGAGATGACCTGCAATTCGCTTGGTGTACTCTG 1527
Db 241 AsnAsnThrArgGlnGlnThrGluAlaArgAspAspLeuHisCysProTIPCysThrLeu 260
QY 1528 AACTCCGCAAACTTTATGATTACTCAAGCATCTTAACATCTGCCATAGCAGATTTATC 1587
Db 261 AsnCysArgLysLeuYsthrSerLeuLeuLysHisLeuLysLeuCysHisSerArgPheIle 280
QY 1588 TTCAACTATGTTATC---ATCCAAAGGTGCTAGGA 1621
Db 281 PheAsn***Val-LeuSerSerGlnLysValLeuGly 292
```

RESULT 6

```
US-09-764-864-1355
; Sequence 1355, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1355
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
```

```
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (166)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1355
```

```
Alignment Scores:
Pred. No.: 4,2e-73 Length: 175
Score: 940.00 Matches: 172
Percent Similarity: 97.73% Conservative: 0
Best Local Similarity: 97.73% Mismatches: 3
Query Match: 12.22% Indels: 1
DB: 9 Gaps: 0
```

US-09-874-162A-7 (1-4409) x US-09-764-864-1355 (1-175)

```
QY 1447 TTTTATCAGTTTCTCTATTAACAACTAACAAGCAAACTGAAGCAAGAGATGACCTG 1506
Db 1 PheTyrGln***LeuTyrAsnAsnThrArgGlnGlnThrGluAlaArgAspLeu 20
QY 1507 CATTGCCCTTGGTGTACTCTGAACCTGCCCAACTTTATAGTTTACTCAAGCATCTTAAA 1566
Db 21 HisCysProTIPCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLys 40
QY 1567 CTCTGCCATAGCAGATTATTCATTCACATGTTTATCATCCAAAAGGTGCTAGGATAGAT 1626
Db 41 LeuCysHisSerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAsp 60
QY 1627 GTTCTCTAATCAGTGTATTATGATCGCTTCATGAGGAATCTCAGATATTCATCGC 1686
Db 61 ValSerIleAsnGluCysTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArg 80
QY 1687 CAACCTGGATTTGCTTTTAGTCGCAACCGACCACTTAAGAGAACACCTATCACACATAT 1746
Db 81 GlnProGlyPheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIle 100
QY 1747 CTGTGTGAGGCCAAACAGAAACAGCAACAGCATGCTCTGAATTTCTTGAATCTGAAGAT 1806
Db 101 LeuValCysArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAsp 120
QY 1807 GGGGAAGTAGAAGCAGCAAAAGACATATAGTAGTGGCCCAATCGTCTGTATTTCATAGT 1866
Db 121 GlyGluValGluGlnGlnArgThrTyrSerSerGlyHisAsnArg***TyrPheHisSer 140
QY 1867 GATACCTGTCTTACCTCTCCGTCACCAAGAAATGGAAGTAGATAGTAGAAGTGAAGAT 1926
Db 141 AspThrCysLeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAsp 160
QY 1927 CTTGAATGGCTAAGAGAAAACCATTTACACAAATTTGAAGAGTTTT 1972
Db 161 ProGluTIPLeuArg-***LysProLeuHisLysLeuLysSerPhe 175
```

RESULT 7

```
US-09-764-864-1334
; Sequence 1334, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1334
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
```


; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-874-162A-7 (1-4409) x US-09-874-162A-2 (1-243)

Alignment Scores:
Pred. No.: 6,37e-55 Length: 278
Score: 733.00 Matches: 142
Percent Similarity: 98.62% Conservative: 1
Best Local Similarity: 97.93% Mismatches: 2
Query Match: 9.53% Indels: 0
DB: Gaps: 0

US-09-874-162A-7 (1-4409) x US-09-764-864-1334 (1-278)

QY 1 CCGCCCGCGCTCGCAGAGCCGACACACCGGGGCTCTCGATGTAGCAACATGACAGGC 60
DB |||||:|||||
QY 19 ProAlaArgSerGlnAsnArgHisGlnGlySerGlyCysSerThrMetThrGly 38
DB |||||:|||||
QY 61 ATCGCGCGCGCTCTCTTCTCCCAATACCTCCGATTCGGGGGCTCGGACTCCCACTTC 120
DB |||||:|||||
QY 39 IleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGlyLeuHisPhe 58
DB |||||:|||||
QY 121 CCCACCTCGCGACTCATCGACACATCGAGGCAACCCATCGATACAGATCCAGG 180
DB |||||:|||||
QY 59 ProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThrAspProArg 78
DB |||||:|||||
QY 181 GTTTTAGAAAAACAAGATTACAGCAGCAACCTATGTTGCCCTGAGTTACATAATAGA 240
DB |||||:|||||
QY 79 ValLeuGluLysGlnGluLeuGlnProThrTyrValAlaLeuSerTyrIleAsnArg 98
DB |||||:|||||
QY 241 TTCATGACAGATGCTGCCCGCGAGCAGGAGTCCCTAAAGAAGATTTCAGCCGAAG 300
DB |||||:|||||
QY 99 PheMetThrAspAlaAlaArgGluGlnGluSerLeuLysLysLysIleGlnProLys 118
DB |||||:|||||
QY 301 CTCCTCGTGACTCTGCAGCTCAGTCTCGAGGGAATGTCTCCCTCCCGCCAGCCAC 360
DB |||||:|||||
QY 119 LeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrProProArgHis 138
DB |||||:|||||
QY 361 AGCAGTGAAGCTTACTCCCGCGTACCCCGTACCCCGCCATCCCGCTCTCTTCATTCGCG 420
DB |||||:|||||
QY 139 SerSerGlySerLeuThr***ProValThrProProIleThrProSerSerSerPheArg 158
DB |||||:|||||
QY 421 ACCAGCATCCGACA 435
DB |||||:|||||
QY 159 SerSerThrProThr 163

RESULT 8
US-09-874-162A-2
; Sequence 2, Application US/09874162A
; Patent No. US20020155452A1
; GENERAL INFORMATION:
; APPLICANT: Koontz, Jason
; TITLE OF INVENTION: FUSION OF JAZF1 AND JUJAZ1 GENES IN
; FILE REFERENCE: 05311-024001
; CURRENT APPLICATION NUMBER: US/09/874,162A
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,093
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-162A-2
Alignment Scores:
Pred. No.: 1,21e-48 Length: 243
Score: 661.00 Matches: 128
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.60% Indels: 0
DB: Gaps: 0
US-09-874-162A-7 (1-4409) x US-09-874-162A-2 (1-243)

QY 52 ATGACAGCATCGCGCGGCTCTCTTCTTCCAAATACCTGCCGATTCGGGGGCTCGGGA 111
DB |||||:|||||
QY 1 MetThrGlyIleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGly 20
DB |||||:|||||
QY 112 CTCCACTTCCCGCCCGCTCTCATCGAGCACATCGAGGACACCAACCATCGATACA 171
DB |||||:|||||
QY 21 LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThr 40
DB |||||:|||||
QY 172 GATCCACGGGTTTAGAAAAACAAGATTACAGCAGCAACCTATGTTGCCCTGAGTTAC 231
DB |||||:|||||
QY 41 AspProArgValLeuGluLysGlnGluLeuGlnProThrTyrValAlaLeuSerTyr 60
DB |||||:|||||
QY 232 ATAAATAGATTATGACAGATGCTGCCCGCGAGCAGGAGTCCCTTAAAGAGAGAGATT 291
DB |||||:|||||
QY 61 IleAsnArgPheMetThrAspAlaAlaArgGluGlnGluSerLeuLysLysLysIle 80
DB |||||:|||||
QY 292 CAGCGAAGCTCTGCTGACTCTGTCCAGTCTCGAGGGAATGTGTCCACTCCC 351
DB |||||:|||||
QY 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100
DB |||||:|||||
QY 352 CCAGCCGACAGCAGTGAAGCTTACTCCCGCGTACCCCGCCATCCCGCTCTCTCT 411
DB |||||:|||||
QY 101 ProArgHisSerSerGlySerLeuThrProProValThrProIleThrProSerSer 120
DB |||||:|||||
QY 412 TCATTCCGCGCAGCAGCTCCGACA 435
DB |||||:|||||
QY 121 SerPheArgSerSerThrProThr 128

RESULT 9
US-10-144-198-16
; Sequence 16, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 90 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-16
Alignment Scores:
Pred. No.: 1,21e-48 Length: 243
Score: 661.00 Matches: 128
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.60% Indels: 0
DB: Gaps: 15
US-09-874-162A-7 (1-4409) x US-10-144-198-16 (1-243)

QY 52 ATGACAGCATCGCGCGGCTCTCTTCTTCCAAATACCTGCCGATTCGGGGGCTCGGGA 111
DB |||||:|||||
QY 1 MetThrGlyIleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGly 20
DB |||||:|||||
QY 112 CTCCACTTCCCGCCCGCTCTCATCGAGCACATCGAGGACACCAACCATCGATACA 171
DB |||||:|||||
QY 21 LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThr 40
DB |||||:|||||
QY 172 GATCCACGGGTTTAGAAAAACAAGATTACAGCAGCAACCTATGTTGCCCTGAGTTAC 231
DB |||||:|||||
QY 41 AspProArgValLeuGluLysGlnGluLeuGlnProThrTyrValAlaLeuSerTyr 60
DB |||||:|||||

```

QY 232 ATAAATAGATTTCATGACAGATGCTCCCGCGAGAGCAGGAGTCCCTAAAGAAAGAGATT 291
Db 61 IleaAsnArgPheMetThrAspAlaAlaArgGlnGlnGluSerLeuLysLysLysLe 80
QY 292 CAGCCGAAGCTCTCGCTGACTCTCTCCAGCTCAGTGTCTCGAGGGAATGTCCACTCCC 351
Db 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100
QY 352 CAGCCGACACAGTGGAGCTTACTCCCGCTGACCCCGACCCACCATCACCCTCTCT 411
Db 101 ProArgHisSerSerGlySerLeuThrProProValThrProProIleThrProSer 120
QY 412 TCATTCCCGCAGCAGCTCCGACA 435
Db 121 SerPheArgSerSerThrProThr 128

RESULT 10
US-10-104-047-2404
; Sequence 2404, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent in Ver. 2.1
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2404

Alignment Scores:
Pred. No.: 1,21e-48 Length: 243
Score: 661.00 Matches: 128
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.60% Indels: 0
Db: 15 Gaps: 0

US-09-874-162A-7 (1-4409) x US-10-104-047-2404 (1-243)

QY 52 ATGACAGCATCGCGCGCTCTCTTCTCCATACCTCCGAGTCCGGGGTCCGGA 111
Db 1 MetThrGlyIleAlaAlaSerPheSerAsnThrCysArgPheGlyGlyCysGly 20
QY 112 CTCACCTTCCACCTCGCGACCTCATCGAGCAGCATCGAGGACCAACCATCGATACA 171
Db 21 LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisLeAspThr 40
QY 172 GATCCACGGGTTTATAGAAAAACAAGATTACAGCAGCCCAACTATGTCCTCGATTAC 231
Db 41 AspProArgValLeuGluLysGlnGlnGlnProThrTyrValAlaLeuSerTyr 60
QY 232 ATAAATAGATTTCATGACAGATGCTCCCGCGAGAGCAGGAGTCCCTAAAGAAAGAGATT 291
Db 61 IleaAsnArgPheMetThrAspAlaAlaArgGlnGlnGluSerLeuLysLysLysLe 80
QY 292 CAGCCGAAGCTCTCGCTGACTCTCTCCAGCTCAGTGTCTCGAGGGAATGTCCACTCCC 351
Db 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100
QY 352 CCACGCCACAGCAGTGGAGCTTACTCCCGCTGACCCCGACCCACCATCACCCTCTCT 411
Db 101 ProArgHisSerSerGlySerLeuThrProProValThrProProIleThrProSer 120
QY 412 TCATTCCCGCAGCAGCTCCGACA 435

```

```

Db 121 SerPheArgSerSerThrProThr 128

RESULT 11
US-09-764-864-884
; Sequence 884, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 884
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-884

Alignment Scores:
Pred. No.: 3,05e-40 Length: 143
Score: 564.00 Matches: 109
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.09% Mismatches: 0
Query Match: 7.33% Indels: 0
Db: 9 Gaps: 0

US-09-874-162A-7 (1-4409) x US-09-764-864-884 (1-143)

QY 1 CCGCGCGCGCTCCGAGCGCAGCACCGAGGGGCTCTCGATGTAGCACCATGACAGGC 60
Db 34 ProAlaArgArgSerGlnAsnArgHisGlnGlyGlySerArgCysSerThrMetThrGly 53
QY 61 ATCGCGCGCGCTCTCTTCTTCCAAATACCTGCGGAGTCCGAGTCCGAGTCCACTTC 120
Db 54 IleaAlaAlaSerPheSerAsnThrCysArgPheGlyGlyCysGlyLeuHisPhe 73
QY 121 CCGACCTGCGGAGCTCATCGAGCAGCATCGAGGACACCATGATACAGATCCACGG 180
Db 74 ProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisLeAspThrAspProArg 93
QY 181 GTTTTAGAAAAACAAGATTACAGCAGCCCAACTATGTCCTGAGTTACATAAATAGA 240
Db 94 ValLeuGluLysGlnGlnGlnProThrTyrValAlaLeuSerTyrIleAsnArg 113
QY 241 TTCATGACAGATGCTCCCGCGAGCAGGAGTCCCTAAAGAAAGAGATTCCAGCGAG 300
Db 114 PheMetThrAspAlaAlaArgGlnGlnGluSerLeuLysLysLysLeuGlnProLys 133
QY 301 CTCCTGCTGCTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 330
Db 134 LeuSerLeuThrLeuSerSerSerValSer 143

RESULT 12
US-10-029-386-28495
; Sequence 28495, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28495
; LENGTH: 69
; TYPE: PRT

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR17.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P23477, EVALUE 2.70e+00
US-10-029-386-28495

Alignment Scores:
Pred. No.: 2,01e-22 Length: 69
Score: 359.00 Matches: 68
Percent Similarity: 98.55% Conservative: 0
Best Local Similarity: 98.55% Mismatches: 1
Query Match: 4.67% Indels: 0
DB: 14 Gaps: 0

US-09-874-162A-7 (1-4409) x US-10-029-386-28495 (1-69)
QY 1756 AGCCCAAAAGCAAGCAAGCATGCTGATTTCTTGAATCTTGAATCGAAGTGGGAAGTA 1815
Db 1 ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal 20
QY 1816 GAACAGCAAAAGCAATATAGTAGTGCCCAATCGTCTGTATTTCCATAGTAGTACCTGC 1875
Db 21 GluGlnArgThr-Tyr-SerSerGlyHisAsnArgLeuTyr-PheHisSerAspThrCys 40
QY 1876 TTACTCTTCGTCACAGAAATGCAAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 1935
Db 41 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAspProGluTrp 60
QY 1936 CTAAGAGAAAAAACCAATTACACAAATT 1962
Db 61 LeuArgGluLysThrIleThrValIle 69

RESULT 13
US-10-276-774-2060
; Sequence 2060, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2060
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-2060

Alignment Scores:
Pred. No.: 2,27e-15 Length: 171
Score: 280.00 Matches: 56
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.64% Indels: 0
DB: 12 Gaps: 0

US-09-874-162A-7 (1-4409) x US-10-276-774-2060 (1-171)
QY 268 CAGGAGTCCCTAAAGCAAGATTGAGCGGAGCTCTCGCTGCTCTGCTCTGCTCTGCT 327
Db 1 GlnGluSerLeuLysLysLysIleGlnProLysLeuSerLeuThrLeuSerSerVal 20

QY 328 TCTCAGGGAATGTCTCACTTCCCTCCAGCCACAGCAGTGGAGCCTTACTTCCCCCGTG 387
Db 21 SerArgGlyAsnValSerThrProArgHisSerGlySerLeuThrProProVal 40
QY 388 ACCCCACCCATCACCCCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 435
Db 41 ThrProIleThrProSerSerSerPheArgSerSerThrProThr 56

RESULT 14
US-10-231-778-2
; Sequence 2, Application US/10231778
; Publication No. US20030126647A1
; GENERAL INFORMATION:
; APPLICANT: Bilodeau, Pierre
; APPLICANT: Chaudhury, Abdul M.
; APPLICANT: Dennis, Elizabeth S.
; APPLICANT: Koltunow, Anna M.G.
; APPLICANT: Luo, Ming
; APPLICANT: Peacock, William J.
; TITLE OF INVENTION: Method for inducing seed development by down-regulating
; TITLE OF INVENTION: expression of the FIS2 gene
; FILE REFERENCE: 72-98A
; CURRENT APPLICATION NUMBER: US/10/231,778
; CURRENT FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: 09/398,237
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 60/101,184
; PRIOR FILING DATE: 1998-09-21
; PRIOR APPLICATION NUMBER: AU PP6061
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: AU PP6062
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: AU PP6063
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: AU P01345
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: AU P01346
; PRIOR FILING DATE: 1999-07-01
; NUMBER OF SEQ ID NOS: 239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-231-778-2

Alignment Scores:
Pred. No.: 2,14e-09 Length: 813
Score: 214.50 Matches: 147
Percent Similarity: 32.70% Conservative: 113
Best Local Similarity: 18.49% Mismatches: 298
Query Match: 2.79% Indels: 237
DB: 14 Gaps: 28

US-09-874-162A-7 (1-4409) x US-10-231-778-2 (1-813)
QY 436 GAGCCCAACACAGATCTATGATTTCTTGAATCTTGAATCTTGAATCTTGAATCTTGAAT 495
Db 70 LysProValAlaValTyrLysLysLeuGluThrArgSerLysAsnAsnProTyrPheLeu 89
QY 496 CACAGAACTTCTTACTACATG-----TCTCATCGAACTCCAGAACATCAAAAGG 549
Db 90 ArgArgSerLeuLysTyrIleIleGlnAlaLysLysLysLysSerAsnSerGlyGly 109
QY 550 AAA---ACATTAAAGTTGATGATATGTTATCAAAAGTAGAGAAAATGAAAGGAGAGCAA 606
Db 110 LysIleArgPheAsnTyrArgAspValSerAsnLysMetThrLeuLysAlaGluValVal 129
QY 607 GAATCTCATGAC-----TTGTCAGCT 627
Db 130 GluAsnPheSerCysProPheCysLeuIleProCysGlyGlyHisGluGlyLeuGlnLeu 149

```

QY 628 CATTG----- 633
 Db 150 HisLeuLysSerSerHisAspAlaPheLysPheGluPheTyrArgAlaGluLysAspHis 169
 QY 633 ----- 633
 Db 170 GlyProGluValAspValSerValLysSerAspThrIleLysPheGlyValLeuLysAsp 189
 QY 634 ----- 634
 Db 190 AspValGlyAsnProGlnLeuSerProLeuThrPheCysSerLysAsnArgAsnGlnArg 209
 QY 679 CCAAACTCAGAAATGAAACAAATTCCTGTTACCTCGGAAGTCCTGCTGTGCAAGTT--- 735
 Db 210 ArgGlnArgAspAspSerAsnAsnValLysLysLeuAsnValLeuLeuMetGluLeuAsp 229
 QY 736 ----- 736
 Db 230 LeuAspAspLeuProArgGlyThrGluAsnAspSerThrHisValAsnAspAspVal 249
 QY 760 AGTTGTCCTCAATAAGG-----CAA 777
 Db 250 SerSerProProArgAlaHisSerSerGluLysIleSerAspIleLeuThrThrGln 269
 QY 778 GTTCCACAGGTAAAGACAGCGTGCCTTTGATTCCTGACCTCAATCAACAAACCCGGA 837
 Db 270 LeuAlaIleAlaGluSerSerGluProLysValProHisValAsnAsp-----Gly 286
 QY 838 AATTTC-----CCGTCCCTTGCAGTTCCAGTAATGAA----- 870
 Db 287 AsnValSerSerProProArgAlaHisSerSerAlaGluLysAsnGluSerThrHisVal 306
 QY 871 -----TTTGAACCTAGTAACGCCATATGTTGGAAGCTTTACTCGTTG 912
 Db 307 AsnAspAspAspValSerSerProProArgAlaHisSerLeuGluLysAsnGluSer 326
 QY 913 CTATTT-----AGAGTCACTCGTCCAGGAAGAGAGAGTTAATGATGAT 960
 Db 327 ThrHisValAsnGluAspAsnIleSerSerProProLysAlaHis----- 341
 QY 961 AATGAGAAACAATGAAATATTGATGCTCAATGAAGAG-----CTTCCAGCCAGA 1011
 Db 342 SerSerLysLysAsnGluSerThrHisMetAsnAspGluAspValSerPheProArg 361
 QY 1012 AGAAACGAATTCGTGAGATGGGGAAGACATTTGTTGCAAAATGACAGATTTGAT 1071
 Db 362 ThrArgSerSerLysGluThrSerAspIleLeuThrThrGlnProAlaIleValGlu 381
 QY 1072 AAA-----AACAGGCGCTTACAGCTTTTAGATGGGAATAT 1107
 Db 382 ProSerGluProLysValArgArgGlySerArgArgLysGlnLeuTyrAlaLysArgTyr 401
 QY 1108 GAAGTAGCCATGCAGGAATCGAAGAATGTCCATAAGCAAGAAAGCAACATGGGAG 1167
 Db 402 -----LysAlaArgGluThrGlnPro 408
 QY 1168 ACTATTCTTGATGGAGAGAGCTGCCTCCATTC-----GAAACATTTTCTCAG 1215
 Db 409 AlaIleAlaGluSerSerGluProLysValLeuHisValAsnAspGluAsnValSerSer 428
 QY 1216 GGACCT-----ACGTTGCACTTCCACT 1236
 Db 429 ProProGluAlaHisSerLeuGluLysAlaSerAspIleLeuThrThrGlnProAla 448
 QY 1237 CTTCGTTGCAGAGAG-----ACCAATGTAATCTACGGCTCCTATT 1281
 Db 449 IleAlaGluSerSerGluProLysValProHisValAsnAspGluAsnValSerSerThr 468
 QY 1282 GCCAAACCTCTGCCACTAGAAATTCAGAGAGTCTCCATCAG-----GAAACCAAGCCT 1335
 Db 469 ProArgAlaHisSerSerLysAsnLysSerThrArgLysAsnValAspAsnValPro 488
 QY 1336 GGTTCAGTTAAACCTACTCAAACTATTCTCTGTTAAAGATCAATTCAGTACAGTCTCAAA 1395

Db 489 SerProProLysThrArgSerSerLysLysThrSerAspIleLeuThrThrGlnPro 508
 QY 1396 ACAAGAAAGAAAGAGATACCTCA-----AATGAAACCGACAAAA 1437
 Db 509 ThrIleAlaGluSerSerGluProLysValArgHisValAsnAspAspValSerSer 528
 QY 1438 TTAAGAAATATTTATCAGTTCTCTATTAACAACATCAACGCAACAACTGAAGCAAGA 1497
 Db 529 ThrProArgAlaHisSerSerLysLysAsnLysSerThrArgLysAsnAsp----- 545
 QY 1498 GATGACCTGCATGCTTGGTGTACTCTGAACCTGCCCAACTTATAGTTTACTCAAG 1557
 Db 546 AspAsnIleProSerProLysThrArgSerSerLysLysThrSerAsnIleLeuThr 565
 QY 1558 CATCTTAACCTGCGCATGACAGATTTATCTCAACTATGTTATCATCCAAAAGGTGCT 1617
 Db 566 ArgThrGln-----ProAlaIleAla 572
 QY 1618 AGGATAGATGTTTCTATCAATCAGTGTATGATGCTCTCTATGACGAGAAATCCT----- 1671
 Db 573 GluSerGluProLysValProHisValAsnAspLysValSerSerThrProArgAla 592
 QY 1672 -----CAGGATATTCATCGCAACCTGGATTGCT----- 1701
 Db 593 HisSerSerLysLysAsnLysSerThrHisLysLysAspAspAsnAlaSerLeuProPro 612
 QY 1702 -----TTTAGTCGCAACGACCACTAGTAAAGAGA 1728
 Db 613 LysThrArgSerSerLysLysThrSerAspIleLeuAlaThrThrGlnProAlaLysAla 632
 QY 1729 ACACCTATCACACATATTCCTGTGTGAGGCGCAACGAAACGAAAGCAAGCATGCTGAA 1788
 Db 633 GluPro-----SerGluProLysValThrArgValSerArgArgLys 646
 QY 1789 TTTCTTGAATCTGAAGATGGGAGAGTGAAGACAGCAAGAAAGATATAGTAGTGGCCACAAT 1848
 Db 647 GluLeuHisAlaGluArgCysGluAlaLysArgLeuGluArgLeuLys-----Gly 663
 QY 1849 CTTCTGTATTTCATGATGATACCTGCTTACTCTCTCCCTCCACAGAAATG-----GAA 1902
 Db 664 ArgGlnPheTyrHisSerGlnThrMetGlnProMetThrPheGluGlnValMetSerAsn 683
 QY 1903 GTAGATAGTGAAGATGAAAGGATCCTGAATGGCTTAAGAGAAACCAATACACAAAT 1962
 Db 684 GluAspSerGluAsnGluThrAspAspTyrAlaLeuAspIleSerGluArgLeuArgLeu 703
 QY 1963 GAAGAGTTTCTGATGTTAATGAAGAGAGAGAAAGATGATGAAGAACTCTGAATCTCCAT 2022
 Db 704 GluArgLeuValGlyValSerLysGluLysArgTyrMetTyrLeuTyrAsnIlePhe 723
 QY 2023 GTTCATGACCATGGTATTATTCCTGACCAATCAAATGAATCATGCTGCTGATGCTGTTGTA 2082
 Db 724 ValArgLysGlnArgValIleAlaAspGlyHisValProTyrPalaCysGluGluPheAla 743
 QY 2083 GAA-----AATTATGACAGAAATAAT 2106
 Db 744 LysLeuHisLysGluGluMetLysAsnSerSerSerPheAspTyrTyrArgMetPhe 763
 QY 2107 AAGACGAATTTATGTCGAAACTTCATCTGCTTCTATCTAGTCAGCATGCTGACTTAACTCT 2166
 Db 764 ArgIleLysLeuTyrAsnAsnGlyLeuIleCysAlaLysThrPheHisLysCysThrThr 783
 QY 2167 ATTGACATATCTCAATAGATAAAGCTGTACCAAGCTCCGTGAATGCAGCAAAAAATTA 2226
 Db 784 IleLeuLeuSerAsnSerAspGluAla-----GlyGlnPhe 795
 QY 2227 GAAAGGGGGAATCTGCTTCCCTCCAAACGAAAGAAATACTGAA 2271
 Db 796 ThrSerGlySerAlaAlaAsnAlaAsnAsnGlnGlnSerMetGlu 810
 RESULT 15
 US-10-177-478-6

```

; Sequence 6, Application US/10177478
; Publication No. US20030165903A1
; GENERAL INFORMATION:
; APPLICANT: Dang, Van-Dinh
; APPLICANT: Okamuro, Jack
; TITLE OF INVENTION: Chimeric Histone Acetyltransferase
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 11696-037801
; CURRENT APPLICATION NUMBER: US/10/177,478
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: U.S. 60/300,135
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 692
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-10-177-478-6

```

Alignment Scores:

```

Pred. No.: 4,09e-09 Length: 692
Score: 211.00 Matches: 149
Percent Similarity: 32.45% Conservative: 122
Best Local Similarity: 17.84% Mismatches: 286
Query Match: 2.74% Indels: 30
DB: 14 Gaps: 30

```

US-09-874-162A-7 (1-4409) x US-10-177-478-6 (1-692)

```

QY 106 TCGGACCTCCACTTCCCGACCTCATCGACGACATCGAGGACCAACAC--- 162
DB 20 CysGlyGlyHis-----GluGlyLeuGlnLeuHisLeuLysSerSerHisAsp 35
QY 163 -----ATCGATACAGATCCACGGGTT----- 183
DB 36 AlaPheLysPheGluPheTyArgAlaGluLysAspHisGlyProGluValAspValSer 55
QY 184 -----TTAGAAAACAGATATACACGCCACCTAT 216
DB 56 ValLysSerAspThrIleLysPheGlyValLeuLysAspValGlyAsnProGlnLeu 75
QY 217 GTTGCCTCGAGTACATAAATAGATTTCATGACAGATGTCGCCGCGAGAGCAGGATCC 276
DB 76 SerProLeuThrPheCysSerLysAsnArgAsnGlnArgGlnArgAspAspSerAsn 95
QY 277 CTAAGAGAGAGATTCAG-----CCGAGGCTC 303
DB 96 AsnValLysLysLeuAsnValLeuLeuMetGluLeuAspLeuAspLeuProArgGly 115
QY 304 TCGCTGACTCTCTCCAGTCTCGAGGATGTCTCCTCCCGCCAGC---CAC 360
DB 116 ThrGluAsnAspSerThrHisValAsnAspAspAsnValSerProProArgAlaHis 135
QY 361 AGCACTGGAAGCTTTACTCCCGCTGACCCACCCATCACCCTCTCTTCATTCGCG 420
DB 136 SerSerGluLysIleSerAspIleLeuThr-----ThrThrGlnLeuAlaAlaGlu 153
QY 421 AGCAGACTCCGACAGACACAGATCTATAGATTCTTCGAACTCGGAATCTATA 480
DB 154 SerSerGluProLysValPro-----HisValAsnAspGlyAsnValSer 168
QY 481 GCACCAATATTTTGCACAGAACTCTTACTATCATGTCTCATGAACTCCAGACAAAC 540
DB 169 SerProProArgAlaHisSerSer-----AlaGluLysAsnGluSerThrHis 184
QY 541 ATCAAAAGGAAACATTTAAAGTTCATGATATGTTATCAAAAGTAGAAATGAAGGA 600
DB 185 ValAsn-----AspAsp----- 193
QY 601 GAGCAAGATCTCATAGCTTGCACCTCATTTGCGAGCTTACGTTTACTGGTTTCTTCCAC 660
DB 194 ProProArgAlaHisSerLeu-----Glu 201

```

```

QY 661 AAAAATGATAAGCATCAACCAAACTCAGAAAAT----- 693
DB 202 LysAsnGluSerThrHisValAsnGluAspAsnIleSerSerProProLysAlaHisSer 221
QY 694 ---GAACAAAATTCCTGTACCTCGAAGTCCTCTGTGAAAGTTTGCACAAAAGAA 750
DB 222 SerLysLysAsnGluSerThr-----HisMetAsnAsp 232
QY 751 AAGGATGTAAGTTGTCATAAAG-----CAAGTTCCCA 786
DB 233 GluAspValSerPheProProArgThrArgSerSerLysGluThrSerAspIleLeuThr 252
QY 787 GGTAAAGACGAGGTGCTTTGATTCCTGACCTCAATCAAAACAAACCCGAAAT----- 840
DB 253 ThrThrGlnProAlaIleValGluProSerGluProLysValArgGlySerArgArg 272
QY 841 -----TTCCTGCTCCCTTGCAGTTTCC 861
DB 273 LysGlnLeuTyralLysArgTyLysAlaArgGluThrGlnProAlaIleAlaGluSer 292
QY 862 AGT-----AATGAATTTGAACCTAGTAACAGCAT 891
DB 293 SerGluProLysValLeuHisValAsnAspGluAsnValSerSerProProGluAlaHis 312
QY 892 ATGCTGAAGTCTTACTCGTTGCTATTATTAGAGTCACTCGTCCAGGA----- 936
DB 313 SerLeuGluLysAlaSerAspIleLeuThrThrGlnProAlaIleAlaGluSerSer 332
QY 937 -----AGAAGAGAGTTTAAT 951
DB 333 GluProLysValProHisValAsnAspGluAsnValSerSerThrProArgAlaHisSer 352
QY 952 GGAATGATTAAATGAGAAACCAATGAAATATTGATGTCATGAAGAGCTTCCAGCCAGA 1011
DB 353 SerLysLysAsnLysSerThrArgLysAsnValAsp---AsnValProSerProProLys 371
QY 1012 AGAAACGAAATCGTGAGGATGGGAAAGACATTTGTCACAAATACAGATTTTGAT 1071
DB 372 ThrArgSerSerLysLysThrSerAspIleLeuThrThrGlnProThrIleAlaGlu 391
QY 1072 AAAACAGGGCTTACAGCTTTTAGTGGGAATATGATGAGTACCATGAGGAAATGGAA 1131
DB 392 SerSer-----GluProLysValArgHisValAsn 401
QY 1132 GAATGCTCAATAAGCAAGAAAGAGCAACATGGGAGACTATTCTTGATGGGAAGGCTG 1191
DB 402 AspAspAsnValSerSer----- 407
QY 1192 CTCTCATTCGAAACATTTTCTCAGGACCTACGTTGCAGTTCACTCTTCTGTCAGCAGA 1251
DB 408 -----ThrProArgAlaHisSer 413
QY 1252 GAGACCAATGATAAATCTACGGCTCTATTGCCCCAACCTCTTGCCCACTAGAAATTCAGAG 1311
DB 414 SerLysLysAsnLysSerThr-----ArgLysAsnAspAsp 425
QY 1312 AGTCTCCATCAGAAACAGCCTGTTGAGTTAAACCTACTCAAACTATTGCTGTATA 1371
DB 426 AsnIleProSerProProLysThrArgSerSerLysLysThrSerAsnIleLeuThrArg 445
QY 1372 GAATCATTCAGTACAGATCTACAAACAGAAAGAAAGATACCTCAAAATGAAACCGA 1431
DB 446 -----ThrGlnProAlaIleAlaGluSerGluProLysValProHisValAsnAsp 462
QY 1432 CAAAATTAAGATATTTTATCATGTTTCTTATACAAATACAGAGCAACAACTGAA 1491
DB 463 AspLysValSerSerThrProArg---AlaHisSerSerLysLysAsnLysSerThrHis 481
QY 1492 GCAGAGATGAC---CTGCAATTCCTTGGTGACTCTGACCTGACCTGCGCAAACTTATAGT 1548
DB 482 LysLysAspAspAsnAlaSerLeuProLysThrArgSerSerLysLysThrSerAsp 501

```

```
QY 1549 TTACTCAAGCATCTTAAACTGTGCCATAGCAGATTATCTTCAACTATGTTTATCATCCA 1608
Db      :|||
502 IleLeu----- 503
QY 1609 AAAGGTGCTAGGATAGATGTTTCTATCAATGAGTGTATGATGCTCTATGCAGAAAT 1668
Db      :|||
503 ----- 503
QY 1669 CCTCAGGATATTCATCGCAACCTGGATTGCTTTTGTCCAAACGACGATTAAGAGA 1728
Db      :|||
504 -----AlaThrThrGlnProAlaLysAla 511
QY 1729 ACACCTATCACACATATTCTTGTGTGTCAGGCGCAACGACAAAGCAAGCATGCTGAA 1788
Db      :|||
512 GluPro-----SerGluProLysValThrArgValSerArgArgLys 525
QY 1789 TTTCTTGAATCTGAAGATGGGAAGTAGACACAGCAAGAACATATAGTAGTGGCCACAAT 1848
Db      :|||
526 GluLeuHisAlaGluArgCysGluAlaLysArgLeuGluArgLeuLys-----Gly 542
QY 1849 CGTCTGTATTCCATAGTAGTACCTGCTTACCTCTCGTCCACAGAAATG-----GAA 1902
Db      :|||
543 ArgGlnPheTyrHisSerGlnThrMetGlnProMetThrPheGluGlnValMetSerAsn 562
QY 1903 GTAGATAGTGAAGATCAAAAGATCCTGATGGCTAAGAGAAAAACCATTTACACAAAT 1962
Db      :|||
563 GluAspSerGluAsnGluThrAspAspTyrAlaLeuAspLysSerGluArgLeuArgLeu 582
QY 1963 GAAGAGTTTCTGATGTTTAATGAGGAGAGAGAAAGTAGTGAATCTGGAATCTCCAT 2022
Db      :|||
583 GluArgLeuValGlyValSerLysGluGluLysArgTyrMetTyrLeuTrpAsnIlePhe 602
QY 2023 GTCATGAAGCATGGGTTTATTGCTGACAAATCAATGAATCATGCTGTATGCTGTTTGA 2082
Db      :|||
603 ValArgLysGlnArgValIleAlaAspGlyHisValProTrpAlaCysGluGluPheAla 622
QY 2083 GAA-----AATTATGGACAGAAAAATTAAT 2106
Db      :|||
623 LysLeuHisLysGluGluMetLysAsnSerSerPheAspTrpTrpTrpArgMetPhe 642
QY 2107 AAGAGAATTTATGCGAACTTCATGCTTCTATGTCATGTCATGTCATGCTTTAATCTT 2166
Db      :|||
643 ArgIleLysLeuTrpAsnAsnGlyLeuIleCysAlaLysThrPheHisLysCysThrThr 662
QY 2167 ATTAGCATAATGTCATATAGATAAGCTGTTACCAAGCTCCGCGTGAATGCAGCAAAATTA 2226
Db      :|||
663 IleLeuLeuSerAsnSerAspGluAla-----GlyGlnPhe 674
QY 2227 GAAAGGGGGAATCTGCTTCCCTGCAACGAGAAATACTGAA 2271
Db      :|||
675 ThrSerGlySerAlaAlaAsnAlaAsnGlnGlnSerMetGlu 689
```

Search completed: August 25, 2004, 20:11:22
Job time : 479.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 25, 2004, 18:23:29 ; Search time 138.5 seconds

(without alignments)
6124.310 Million cell updates/sec

Title: US-09-874-162A-7

Perfect score: 7690

Sequence: 1 cccgcgcgcgcgcgcgcag.....tacataaagtccaataatat 4409

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cn2 1/USPTO spool_p/US09874162/runat_25082004_171844_2190/app_query.fasta_1.4551
-DB=PIR_78 -QFMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=BITS -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09874162 @CN_1 1 218 @runat_25082004_171844_2190 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=10 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 188 | 2.4 | 2013 | 2 C71610 | probable membrane |
| 2 | 178.5 | 2.3 | 1314 | 2 T09481 | mating type silenc |
| 3 | 176.5 | 2.3 | 1142 | 2 S59359 | GIN4 protein - yea |
| 4 | 166 | 2.2 | 1507 | 2 T42631 | breast cancer tumo |
| 5 | 161 | 2.1 | 1875 | 2 S39173 | myosin-like protei |
| 6 | 159.5 | 2.1 | 912 | 2 B44450 | ubiquitin-specific |
| 7 | 159.5 | 2.1 | 1263 | 2 T15496 | hypothetical prote |
| 8 | 158 | 2.1 | 1119 | 2 B70126 | surface-located me |
| 9 | 157 | 2.0 | 1208 | 2 T39068 | coiled coil protei |
| 10 | 156 | 2.0 | 1755 | 2 S69838 | TYB protein - yea |
| 11 | 155.5 | 2.0 | 1755 | 2 S57045 | TYB protein - yea |
| 12 | 155.5 | 2.0 | 3724 | 2 T18427 | hypothetical prote |
| 13 | 154 | 2.0 | 1271 | 2 A45555 | glutamate rich pro |
| 14 | 153 | 2.0 | 707 | 2 A46302 | PTB-associated spl |

ALIGNMENTS

RESULT 1

C71610

probable membrane associated protein PFB0615c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: C71610
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
R;Perlea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: C71610
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2013 <GAP>
A;Cross-references: GB:AE001406; GB:AE001362; NID:G3845230; PIDN:AAC71912.1; PID:G384523
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0615C

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 1.08e-05 | Length: | 2013 |
| Score: | 188.00 | Matches: | 236 |
| Percent Similarity: | 33.41% | Conservative: | 183 |
| Best Local Similarity: | 18.82% | Mismatches: | 362 |
| Query Match: | 2.44% | Indels: | 473 |
| DB: | 2 | Gaps: | 60 |

US-09-874-162A-7 (1-4409) x C71610 (1-2013)

| | | | |
|----|-----|---|-----|
| Qy | 688 | GAAATGAACAAATTCCTGTTACCTGGAGTCTCTGTTGAAGTTTCCACAAAAA | 747 |
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 69 | GLUAspCysAsnLysAsnMetThrIleAsnIleMetLeuGlnAsnGluAsnLys | 88 |
| | | | |
| | | | |
| | | | |
| | | | |
| Qy | 748 | AGAAAGATGATAGTTGCTCCAAAGCAAGTCCACACAGTAAAGCAGGTCCTTC | 807 |
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 89 | IleIleLysGluThrCysTyrIleLysAsnValThrAsnLysIleTyrHisThrLeu | 108 |

QY 808 ATTCTGACCTCAATCAACAAACCCGAAATTTCCCGTCCCTTGCAGTTTCCAGTAAT 867
Db : : : : :
109 PheLeuValIleAsnLysHisTyrrHisAsnIleLeuCysSerLeuSerPheGluAsn 128
QY : : : : :
868 GAATTGTAACCTAGTAACAGCATATGCTGAAGTCTTACTCGTTGCTATTTAGAGTCACT 927
Db : : : : :
129 SerPheGluIleLeuAsnThrAsnPheValLysThr 140
QY : : : : :
928 COTCCAGGAGAGAGAGATTAATGGAATGATTAATGAGAA-----ACCAATGAAAT 981
Db : : : : :
141 -----PheLysGlyLysIleLysSerMetAlaCysThrAsnAsnAsn 154
QY : : : : :
982 ATTGATGTCATGAAGACCTCCAGCCAGAGAA-----CGAAATCGTGAG 1029
Db : : : : :
155 IlePheVal-----LeuIleLysLysLysLysLysLysLysLysLysLysLys 171
QY : : : : :
1030 GATGGGAAAGACATTTGTTGCAAAATGACAGTATT-----GATAAAAC 1077
Db : : : : :
172 AsnGlnMetLysSerLysIleLeuAsnGlnAsnValLeuValSerLysHisThrLeuAsp 191
QY : : : : :
1078 AGCGCTTACAGCTTTAGATGGGAATATGAGTAGCCATG-----CAGGAATGAA 1131
Db : : : : :
192 ArgSerLeuLeuMetLysGlyGluAsnAspValIleCysGluSerLysLys 211
QY : : : : :
1132 GAATGTCCTCAATAAGCAAGAAAGACCAATGGGAGACTATTCTTGATGGGAAGGCTG 1191
Db : : : : :
212 GluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 219
QY : : : : :
1192 COTCCATTGCAAAATTTCTCAGGACCTAGCTGTCAGTTCACTCTCTGTTGACACAGGA 1251
Db : : : : :
220 -----ThrAsp 221
QY : : : : :
1252 GAGACCAATGATAATCTACGCTCCTATT---GCCAACTCTTGCACCTAGAAATCA 1308
Db : : : : :
222 AsnLysAsnGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 241
QY : : : : :
1309 GAGCTCTCCATGAGAA-----AACAGCTGTTGAGTAACTCACTCAAACTATT 1362
Db : : : : :
242 GluLysIleAsnGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 261
QY : : : : :
1363 GCTGTTAAGATCATTGACTACAGATCTACAACAGAAAGAAAGAGATACCTCCAAAT 1422
Db : : : : :
262 GluLysLysAsnGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 281
QY : : : : :
1423 GAAACCGACAAAATTAAGAATATTTATCAGTTTCTCTATAACAACATACAGGCAA 1482
Db : : : : :
282 GluLysLysAsnGlu-----AspThrAsnLysAspPro 292
QY : : : : :
1483 CAACTGAGCAGAGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1542
Db : : : : :
293 TyrGluLysLysGluAsnAsp-AsnIleProLeuGlyAspHisSerValGlnTyrAs 312
QY : : : : :
1543 TATAGTTTACTCAAGCATCTTAAACTCTGCCATAGCAGATTTATCTCAACTATGTTT-- 1600
Db : : : : :
312 nIlePheThrPheSerIleLeuAsnLysLysLysLysLysLysLysLysLysLysLysLys 332
QY : : : : :
1601 -----ATCATCCAAAGGTCTAGGATAGATGTTTCTATCAATGAGTGTATGAGTCTC 1656
Db : : : : :
332 rAsnIleLeuProfile-----LysLysMetIleCysPr 345
QY : : : : :
1657 TATGCAAGAAATCCACAGATATTATCCCAACTGCTGATTTGCTTTAGTCGCAACGGA 1716
Db : : : : :
345 oTyraSpGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 365
QY : : : : :
1717 CAGTTAAGAGAACCTTATCACATATTTCTGTGTCGAGCCCAACAGCAACAAAGCA 1776
Db : : : : :
365 nLysAsnAspAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 379
QY : : : : :
1777 AGCATGCTGAAT----- 1789
Db : : : : :
379 uLeuIlePheAsnSerProIleIleThrThrThrTrpIleAspAsnTyrrIlePheLeuI 399
QY : : : : :
1790 -----TTCTTGAAT-----CTGAGATGGGAG----- 1813

Db : : : : :
399 eTyrrPheLeuAsnAsnGluLeuIlePheLeuSerPheAlaLysProCysArgAsnLeuTy 419
QY : : : : :
1814 -----TAAACAGCAAGAAACATATAGTA----- 1837
Db : : : : :
419 rPheTyrrCysIleAsnAsnTyrrSerHisIleThrSerPhePheTyrrLysSerArgAs 439
QY : : : : :
1838 -----GTGGCCACCAATCGTCTGTTATTTCCATAGTGA----- 1870
Db : : : : :
439 nLeuTyrrIleSerPheLysThrLysGluIleValCysPheLysIleArgTyrrTyrrGluI 459
QY : : : : :
1870 ----- 1870
Db : : : : :
459 eProLeuThrValPheLysLysIleGlnThrThrGluGlyAsnTyrrIleAspAlaLysTy 479
QY : : : : :
1871 -----CTGCTTACTCTCCGTCACAGAAATGGAATGGAATGATAGTAA 1914
Db : : : : :
479 rLeuPheArgLysArgProArgTyrrIleAsn-ThrAsnHisAsnGlnSer-AsnAlaLys 498
QY : : : : :
1915 GATGAAAGGAT-----CCTGAATGCTAAGAGAAAAAAC 1950
Db : : : : :
499 AspAspLysAspGlyAsnAspValIleArgGluGluAspPheLeuArgAsnAsnAsn 518
QY : : : : :
1951 ATTACAAATTAAGAGATTTTCTGATGTTAATGAGAGAGAGAAAGAGTATGAAATC 2010
Db : : : : :
519 -----LysAsnPheSerAspValLysLysArg----- 530
QY : : : : :
2011 TGAATCTCCATGTCATGAAGCATGGTTATTGTCGACATCAATGAATCATGCCTGT 2070
Db : : : : :
531 -----AsnAspLysAsn----- 534
QY : : : : :
2071 ATGCTGTTTGTAGAAATTTATGACAGAAATTAATTAAGAAATTTATGCGAACTTC 2130
Db : : : : :
535 -----AsnTyrr-----GluIleIlePheAsnAsnIleLeuArgGluIle 547
QY : : : : :
2131 -----ATGCTTCATCTAGCAGCATGCTTAACTTATCTTATTAGCATATGTCATA 2184
Db : : : : :
548 LysThrLeuGluAsnLysIleSerAsnAsnAspTyrrAsnIlePhe----- 562
QY : : : : :
2185 GATAAAGCTGTTACCAAGCTCCGTCGATGAAATGCAGCAAAATTAGAAAAGGGGAATCTGCT 2244
Db : : : : :
563 -----TyrGluAspGlyGlu----- 567
QY : : : : :
2245 TCCCTGCAACGAGAAATAACTGAAGAAACAAATGGACAGCAA----- 2290
Db : : : : :
568 -----IleAsn-LysAspGluLeuLysAsnArgLeuSerAlaArgSerLeuSerValTy 585
QY : : : : :
2291 -----ATGATTTAGTGAATTAATACTCAAAAGAGAAAGCTTTGGAACA 2334
Db : : : : :
585 rAsnLysTyrrPheAsnLeuAsnLeuGlyHisAsnAsnAsnLysLysLysLysLysTrp----- 603
QY : : : : :
2335 GATAGTCTCTCAGGGGTTTCAAAACAGACAGCAAAACAAAACTCTGAAAAGCTCTAAC 2394
Db : : : : :
604 -----IleArgGlnAspIleArgAsnAsnMet----- 612
QY : : : : :
2395 CCATGTTATGGACAAACACTGAAATTAATACTTTTAGGAATTCATCTCTAAGAAATATGT 2454
Db : : : : :
613 -----TyrHisAsnLysTyrr-----AsnCysValGluGluAspVa 624
QY : : : : :
2455 TTTTGTGTTTAAATCATATGTTCCAAACAGGCACTGTTAGATGAATTAATATTTCAACA 2514
Db : : : : :
624 lCysIleAsnAspTyrrIle-GluLysGluSerIlePheTyrrGluTyrrAspAsnAsnAsn 644
QY : : : : :
2515 AGGATATTTGTATCAGGGTCTCTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2555
Db : : : : :
644 snAspAsnMetLeuTrpSerHisLeuTyrrPheLeuLysLysLysLysLysLysLysPheA 664
QY : : : : :
2556 -----CATTTAC-----ATGTATATCATCTTTTATGATGTCATTA 2589
Db : : : : :
664 spAsnPheHisTyrrAsnAspGluAsnValIleLysLeuLeuAspPheValSerIleIle- 683
QY : : : : :
2590 AAACATCTGTACTTAAAGCAATATTTTCAAGTATTTTAAAGTATTTTAAACTCAACAA 2649

| | | | |
|----|------|---|------|
| QY | 202 | --CAGCAGCAACCTATGTTGCCCTGAGTTACATAAATAGATTTCATGACAGATGCTGCC | 258 |
| Db | 190 | ArgGlnGlnSerLeuSerSerGlyArgIleHisLyshisIleGluSerProPhe | 209 |
| QY | 259 | CGCCGAGCAGGAGTCCCTAAAGAAG-----AGATTCCAGCGGAAGCTTCGCTG | 309 |
| Db | 210 | GlnArgAsnSerLysProValSerLysSerIleValLysAlaProAsnLysValAsnPhe | 229 |
| QY | 310 | ACTCTGCCAGCTCAGTGTCT-----CGAGGGAATGTGTCCACTCCCCACGCCACAGC | 363 |
| Db | 230 | AlaPheGluThrSerLeuAsnAspIleAspLeuAsnLeuLeuThrArgAlaArgAlaSer | 249 |
| QY | 364 | AGTGGAGCCTTACTCCCCCGTACCCACCACCATCACCCCTCCTCTTTCATTCCGCAGC | 423 |
| Db | 250 | SerGlyGly-----ThrProThrSerLysGlySerAla | 260 |
| QY | 424 | AGCATCCGACAGACCCAAACACAGATCTATAGATTCTTCCGAATCCGAAT----- | 474 |
| Db | 261 | AlaValProAlaAlaPro-----LysThrLysAsnLysThrAsp | 273 |
| QY | 475 | -----CTCATAGCCCAATATTTTGCACAGAACTCTTACTTACATGTCAT | 522 |
| Db | 274 | SerGlnValSerLysIleAlaAsnGluIleGlnAsnSerAsnLysSerArgIleSerLys | 293 |
| QY | 523 | CGAACTCCGAACAAACATCAAAGAGAAACATTTAAAGTTGATGATATGTTATCAAAA | 582 |
| Db | 294 | AlaAsnSer-----ProIleGlnArgThrThrThrThrSerAsnIleProSerGlu | 311 |
| QY | 583 | GTAGAGAAATGAAGAGAGCAAGATCTCATAGCTTTCAGCTCATTTTCGACGTTACG | 642 |
| Db | 312 | AlaLysGluLeu-----TyrAsnMetLeuSerSerArgLeuLysArgArg | 326 |
| QY | 643 | TTTACTGGTTCTTCCAAAAATGATAAGCCATCACCAAACTCAGAAAAATGACAAAAAT | 702 |
| Db | 327 | AsnValLeuPheAspLysLysGlyThrIleAsnSerGluAsnAsnAspLysThrGlyAla | 346 |
| QY | 703 | TCTGTTACCTCGGAAGTCCTGCTGTGGAAGTTTGCACAAAAAAGAAAGATGTAAGT | 762 |
| Db | 347 | SerValLysValGluSerSerHisValGluIleIleGlySerLysArgAspHisThrHis | 366 |
| QY | 763 | TGTCCAATAGGCAAGTTCACACAGGTAAAGAGCAGGTGCTTGATTCTCGACCTCAAT | 822 |
| Db | 367 | ThrSerPheMetGluLeuProSerMetGlnAlaThrIleAlaAspLeuAlaLysThrThr | 386 |
| QY | 823 | CAACAAAAACCCGGAATTTCCGTCCTCCGTCAGTTTCCAGTAAATGAATTTGAACCTAGT | 882 |
| Db | 387 | SerLeuValLysGluAsnSerAlaGlnIleGluValAspArgAsnLeuIleGlyProGlu | 406 |
| QY | 883 | AACAGCCATATGGTGAAGTCTTACTCGTTGCTATTTAGATGATCTCCAGGAAGAAGA | 942 |
| Db | 407 | ArgSer-----ArgLysAsnGluLys | 413 |
| QY | 943 | GAGTTTAATGGAATGATTAATGGAGAAACCAATGAAAATATTGATGTCATGAAGAGCTT | 1002 |
| Db | 414 | GluPheGlnAspLeu-----GluThrArgThrSerSerProAlaLysGluSerLeu | 430 |
| QY | 1003 | -----CCAGCCAGAGAAACGAAATCGTAGGATGGGAAAAG | 1041 |
| Db | 431 | AlaSerProSerAlaAspThrProGluArgLysValProGlnIleuAspLeuThrSerPro | 450 |
| QY | 1042 | ACATTTGTCACAAATCACAGTATTTGATAAAAAACAGGCGCTTACAGCTTTTAGATGGG | 1101 |
| Db | 451 | SerPheAspSerThrProGlyValThrGluLysMetAspLeuLysGluValProSerPro | 470 |
| QY | 1102 | GAATATGAA-----GTAGCCATCAGGAAATGGAAGAAATGTCCTCAATTAAG-- | 1146 |
| Db | 471 | GluTyrProSerSerAlaGlyGlnAlaValGlnIleSerGluSerGlnIleSerHis | 490 |
| QY | 1147 | -----AAG | 1149 |
| Db | 491 | ValIleGlnGlnAlaGluGluAlaAspSerGlySerHisGlyIleGlnValGluSer | 510 |

| | | | |
|------|----|---|------|
| 1150 | QY | AAAAGAGCAACATGGGAGACATATCTTGATGGGAAGAGCGCTGCCTCCATTCGAAACATTT | 1209 |
| 511 | Db | SerGluLeuAsnIleProThrIlePheGluHisLysAspAsnPro----- | 525 |
| 1210 | QY | TCTCAGGACCTACGTTGCGATTCCTCTCGTTCGACAGGAGACCAATGATAAATCT | 1269 |
| 526 | Db | ---GluAlaSerAlaSerGlyThrPheHisHisValAspGlu-----LysGlnSer | 542 |
| 1270 | QY | ACGGCTCTTATGCCAAACCTCTTGGCCATCAGAAATTCAGAGATCTCCATCAGGAAAC | 1329 |
| 543 | Db | ThrGluGluThrAlaLysLeuLeuAlaLeuSerAspThrSerSer-----GluThrSer | 560 |
| 1330 | QY | AAGCCTGGTTC-----GTTAAACCTACTCAACTATGCTGTTAAAGAA | 1374 |
| 561 | Db | SerAspGlySerSerSerGlyHisTyrrValGlyAspPheGlnLeuLeuAsnThrMetLys | 580 |
| 1375 | QY | TCATTGACTACAGATCTACAAACAAGAAAGAAAGGATCTCCAATGAAACCGACAA | 1434 |
| 581 | Db | AlaArgMetValGlyProLysValAlaLysTrpLysAlaPhe-----AsnLysSer | 597 |
| 1435 | QY | AAATTAAAGAAATTTATCAGITTCCTCTATACACAATACAAGCAACAACCTGAAGCA | 1494 |
| 598 | Db | LysPheLysAspLeuMetLysTyrrLeuIleAlaGluHisThr-----ProLeuAlaAla | 615 |
| 1495 | QY | AGAGATGACCTGATCGCCTGGTCTACTCTCAACTCGCCAAACTTTATAGTTTAA--- | 1551 |
| 616 | Db | ArgSerAspValSerCysSerAsnAspIleCysAsnIleArgAsnTyrrGluSerValGlu | 635 |
| 1552 | QY | CTCAAGCATCTTAACTCTGCCACACAGCATTTATCTTC----- | 1590 |
| 636 | Db | ValHisHisLeuGlnLysLeuProArgLeuArgTyrrLeuGluLeuGluSerProSer | 655 |
| 1591 | QY | ---AACTATGTTTATCATCCAAAAGTGCT-----AGGATAGAT---GTTTCTATCAAT | 1638 |
| 656 | Db | AsnAsnPheIleSerLysProLeuArgAlaAsnAsnArgValGluLysLeuArgLysAsp | 675 |
| 1639 | QY | GAGTGTATTGATGGCTCTATGCAGGAAATCTCCAGATATTTCATGCCAACTGGATTT | 1698 |
| 676 | Db | GluLeuTyrrAsnGlySer-----LysGlnLysArgArgThrArgGlySer | 690 |
| 1699 | QY | GCTTTTAGTCGCACCGCACCGTTAG----- | 1725 |
| 691 | Db | AlaProAspSerGluGlySerLysLysProLeuIleGluArgGlnIleGluAspGlyAsn | 710 |
| 1726 | QY | -----AGAACCATCTACACATATTCCTGTGTGCAGGGCCAAACGA | 1767 |
| 711 | Db | ValSerGluLysThrProGlnGluProThrThrSerIle-----ProGlyGlu | 726 |
| 1768 | QY | ACAAAAGCAAGCATGCTGAATTTCTTGAATCTGAAGATGGGAAGTAGAACAGCAAGA | 1827 |
| 727 | Db | ThrAspValAsnGluLysValLeuGluProIleGluAsp-----IleGluMetLeuPro | 744 |
| 1828 | QY | ACATATAGTAGTGGCCACCAATCGCTGTGATTTCCATAGTGATACC-----TGC | 1875 |
| 745 | Db | ThrPheIleSerSerSerAsnArgLeuGluProAspAlaAsnThrLeuSerAlaHisAla | 764 |
| 1876 | QY | TTACCT----- | 1881 |
| 765 | Db | LeuProSerSerValIleGlyValLysLysGlnProSerLeuThrThrValAsnLysPhe | 784 |
| 1882 | QY | -----CTCCGTCCACAAGAAATCGAAGTAGATGAA----- | 1914 |
| 785 | Db | ProAlaSerPheGluAsnGlyLeuGlnProAspAsnAlaGlnAlaSerThrGluIleGlu | 804 |
| 1915 | QY | ---GATGAAAAGGATCCTGAATGGCTTAAGAGAAAAACCATTACACAAATTGAAGATTT | 1971 |
| 805 | Db | ThrAsnGluGlyAsnSerLysLysGlyLysAspGluLeuHisPheAspProGluLysCys | 824 |
| 1972 | QY | TCTGATGTTAATGAAGAGAGAAAGATGATCAAACTCTGGAACTCCATGTCATGAG | 2031 |
| 825 | Db | GluAspValThrGlnPheArgLysGluTyrrLeuAlaLeu---ArgAlaSerIleSerLys | 843 |
| 2032 | QY | CATGGGTTTATTGCTCACAAATCAAAATGAATCATGCTGTATGCTGTTCTAGAAATTTAT | 2091 |

```
Db      844 AspThrCysSerSerLeuValAsn----- 853
QY      2092 GGACAGAAATATTAGAGAAATTTATGTCGAAATTCATGCTTCATCTAGTCAGCATG 2151
Db      854 ---GluAspLeuAsnLysThrGluValIleGlnAsnLeuThrSerGluIleIle---Gln 871
QY      2152 CATGATTTAATCTATTAGCATAATGTCAATAGATAAAGCTGTTACCAAGCTCCGTGAA 2211
Db      872 SerGluMetLysIleAlaSerLeuValGlyIleAsnHis-----GlnLeuArgGlu 888
QY      2212 ATCCAGCAAAATAGAAAG-----GGGGAATCTGCTTCCCTGCAACGAAAGATA 2265
Db      889 LysLeuGluGluGluGluLysIleGlnAlaLysLeuValGlnGluLeuAspThrValLeu 908
QY      2266 ACTGAGAAACAAAT-----GGGACAGCAATGGATTAGTGAATTAACATCA 2313
Db      909 LeuGluSerLysGlySerPheSerAlaGlyThrAlaLysGlu---AlaGluValAsnAsn 927
QY      2314 AAAGAGAAAGCTTTGGAACACAGATAGTGTCTCAGGGGTTTCAAAACAGACGCAAAACAA 2373
Db      928 LysAspLys-----ThrGluSerIleAsn-----AsnGluGlnAlaArgGlu--- 941
QY      2374 AAATCTGAAAAGCTTAACCCCATCTGTTATGACAAACACTG 2415
Db      942 -----LysSerLeuThrAlaAlaHisGlyLysLeuIle 952

RESULT 3
S59359
Gln4 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YDR507c
C:Species: Saccharomyces cerevisiae
C:Date: 30-Nov-1995 #sequence_revision 01-Mar-1996 #text_change 24-Sep-1999
A:Accession: S59359; S69565
R:longline, M.S.; Pringle, J.R.
submitted to the EMBL Data Library, August 1995
A:Reference number: S59359
A:Accession: S59359
A:Molecule type: DNA
A:Residues: 1-1142 <LON>
A:Cross-references: EMBL:U33140; NID:g992650; PIDN:AAA75513.1; PID:g992651
R:Dieckrich, F.S.
submitted to the EMBL Data Library, August 1995
A:Description: The sequence of S. cerevisiae cosmid 8166, 9787, 9717, and lambda 3073.
A:Reference number: S69553
A:Accession: S69565
A:Molecule type: DNA
A:Residues: 1-1142 <DIE>
A:Cross-references: EMBL:U33057; NID:g927764; PIDN:AAB64949.1; PID:g927777; MIPS:YDR507C
C:Genetics:
A:Gene: SGD:GIN4
A:Cross-references: SGD:S0002915; MIPS:YDR507C
A:Map position: 4R
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
C:Keywords: ATP
F:17-289/Domain: protein kinase homology <KIN>
F:25-33/Region: protein kinase ATP-binding motif

Alignment Scores:
Pred. No.: 7,02e-05 Length: 1142
Score: 176.50 Matches: 152
Percent Similarity: 34.75% Conservative: 102
Best Local Similarity: 20.79% Mismatches: 264
Query Match: 2.30% Indels: 213
DB: 2 Gaps: 35

US-09-874-162a-7 (1-4409) x S59359 (1-1142)
QY      172 GATCCACGGGTTTATAGAAAACAGAAATTACAGACGCA-----ACC 213
Db      337 AspProGluGlyIle---LysGluLysLeuArgGluProGlyAlaAsnAlaGluLysThr 355
QY      214 TATGTTGCCCTGAGTTACATAAATAGATTTCATGACAGATGTCGCCGCCGAGACGAGAG 273
```

```
Db      356 LeuTyralaLeuLeuTy-----ArgPheLysCysAspThr-----GlnLys 369
QY      274 TCCCTAAAGAGAAGATTACGCCAAGCTTCGCTGACTCTGTCCAGC---TCAGTGTCT 330
Db      370 GluLeuIleLysGlnGlnValLysArgGlnSerIleSerValSerValSer 389
QY      331 CGAGGG---AATGTGTCCACTCCCCCA-----CGCCACACAGTGGAGC 372
Db      390 ProSerLysLysValSerThrThrProGlnArgArgAsnArgGluSerLeuIleSer 409
QY      373 CTTACTCCCCCGTCAGCCCAACCCATC-----ACCCCTCCTCT 411
Db      410 ValThrSerSerArgLysLysProIleSerPheAsnLysPheThrAlaSerSerAlaSer 429
QY      412 TCATTCGCGACGACATCCGACAGACCAACACAGATCTATAGATTCTTCGAACTCGG 471
Db      430 SerSerAsnLeuThrThrProGlySerSerLysArgLeuSerLysAsnPheSerLys 449
QY      472 AATCTCATAGCACCACATATTTTGCACAGAACTTCTTACATGCTCTCATCGAACTCC 531
Db      450 LysLysLeuSerThrIleValAsnGlnSerSerProThrProAlaSer---ArgAsnLys 468
QY      532 AGAACAAACATC-----AAAAGAAACATTTAAAGTTCATGATGATGTTATCAAAAGTA 585
Db      469 ArgAlaSerValIleAsnValGluLysAsnGlnLysArgAlaSerIlePheSerThr 488
QY      586 GAGAAATGAAGAGAGACAGCAAGATCT----- 612
Db      489 LysLysAsnLysArgSerSerArgSerIleLysArgMetSerLeuIleProSerMetLys 508
QY      613 ---CATGATGTGACGCTCATTTGACGTTACGTTTACTGTTTCTTCCACAAATGAT 669
Db      509 ArgGluSerValThrThrLysLeuMetSerThrTyrAlaLysLeuAlaGluAspAsp 528
QY      670 AGCCCATCACCACAACTCAGAAATGAACAAATCTGTTACCTCCCTGGAAGTCCCTGCTG 729
Db      529 TrpGluTyrlleGluLysGluThrLysArgThrSerSerAsnPheAlaThrLeuIleAsp 548
QY      730 AAAGTT-----TGCCACAAAAAAGAGAGATGTAAGTTGT 765
Db      549 GluIlePheGluTyrlleGluLysTyrlleGluLysGluLysGluLysGluLysGlu 568
QY      766 CCAATAGGCAAGTCCACAGT----- 789
Db      569 LysValArgGluAlaLysAlaArgGluGluLysGluLysGluLysGluLysGlu 588
QY      789 ----- 789
Db      589 LysGluArgAlaArgLysLeuLeuLysGluLysGluLysGluLysGluLysGlu 608
QY      790 AAAAAGCAGTGCCTTG---ATTCCTGACCTCAATCAA----- 825
Db      609 LysLysGlnIleGluIleAspIleSerAspLeuGluGlnGluLysLysLysLysGlu 628
QY      826 ---ACAAAACCCGAAATTTCCCGTCCCTGCGAGTTCCAGTAATCAATTTGACCTAGT 882
Db      629 GluLysLeuAspGlyAsnIleArgSerIleSerAlaProMetGluAsnGluLysAsn 648
QY      883 AACAGCCATATGTTGAGTCTTACTCGTTGCTATTAGA-----GTG 924
Db      649 IleAsnHisLeuGluValAspIleAspAsnIleLeuArgArgAsnPheSerLeuGln 668
QY      925 ACTGTCGAGGAGAGAGAGGTTTAATGATGATTAATGGAGAACCAATGAAATATT 984
Db      669 ThrArgProValSerArgLeuAspProGlyIleMetPheSerSerProThrGluGluVal 688
QY      985 GATGTCATGAAGAGCTTCAGCCAGAGAAATCGTGAGGATCGGGAAGACACA 1044
Db      689 SerProValGlu-----ProLysArgThrGluAsnGluArgLeuThrThrGluLys 706
QY      1045 TTTGTTGCCAATG-----ACAGTATTTGATAAAACAGGCGC 1083
```

Db 707 IleLeuGluThrIleArgArgSerLysPheLeuGlySerSerPheAsnIleAspLysGlu 726
 QY 1084 TTACAGCTTTTAGATGGGAATATGAA-----GTAGCCATGCAGGAATG---GAAGAA 1134
 Db 727 LeuLysLeuSerLysMetGluTyProSerIleIleAlaProGlnArgLeuSerGluGlu 746
 QY 1135 TGTCGAATAGCAAGAAAGACACATGGAGACTATCTT-----GATGGAAG 1185
 Db 747 ArgValValSerAspSerAsnAspGlyTyProGluSerLeuIleLeuProLysAspGlyAsn 766
 QY 1186 AGGCTGCCTCCATTCGAA-----ACATTTTCTCAGGACCTAGCTTGCAGTTCACCTCT 1239
 Db 767 GlyValSerGlnLeuLysAspSerThrAlaThrAlaProValSerAspGlyArgLeu 786
 QY 1240 GTTCGACAGAGACCAACGATTAATCTAGGCTCCTATTGCAACCTCTTGCACCT 1299
 Db 787 ArgLysIleSerGlu-----IleArgValProGlnPheThr 798
 QY 1300 AGAATTCAGAGAGCTCCATCAGAAACAAAGCTGTTTCAGTTTAAACCTTACTCAAACT 1359
 Db 799 ArgLysSerArgHisPheSerGluSerAsnLysArgLeuSerVal-----LeuSerMet 816
 QY 1360 APTGCTGTAAAGATCATCTGACTACAGATCTCAAAACAGAAAGAAAGATACCTCCA 1419
 Db 817 TyrSerThrLysGluSerPheThrAsnLeuValAspIleLeuLysAsnGlyAsnLeuAsp 836
 QY 1420 AATGAACACCGACAAATTAAGATATTTTATCAGTTTCTTATACAAACAAATACAAGG 1479
 Db 837 ValAsnAsnGlnGlnSerGlnArgIle----- 845
 QY 1480 CAACAACTGAAGACAGAGATGACCTGCATTGCCCTTGGTGTACTCTGAACCTGCCGAAA 1539
 Db 846 ProThrProArgSerAlaAspAsp----- 853
 QY 1540 CTTTATAGTTTACTCAAGCACTTAACTCTGCCATACAGATTTATCTTCACTATGTT 1599
 Db 854 -----SerGluPheLeuPheGlu----- 859
 QY 1600 TATCATCAAAAGGTGCTAGGATAGATGTTTCTAATGAGTGTATGATGGCTCTCTAT 1659
 Db 860 -----ThrValAsnGlu-----GluAlaGluTyr 867
 QY 1660 GCAGAAATCTCAGATATTCATCGC----- 1686
 Db 868 ThrGlyAsnSerSerAsnAspGluArgLeuTyrAspValGlyAspSerThrIleLysAsp 887
 QY 1687 -----CAACTGGATTGCT-----TTTACTCGAACGGACAGTTAAAGAGA 1728
 Db 888 LysSerAlaLeuLysLeuAsnPheAlaAspArgPheAsnGlySerAsnGluAlaLysGln 907
 QY 1729 ACACCTATCACATATCTTGTGTGCGGCCAAACGAAACAAAGCAAGCATGCTGAA 1788
 Db 908 ThrAspAsnLeuHisLeuProIleLeuProLeuAsnGlyAspAsnGluLeuArgLys 927
 QY 1789 TTTCTTGAATCTGAAGTGGGAGTAGAACAGCAAGCAACATATAGT----- 1836
 Db 928 ---GlnAsnSerGlnGluLysProLysIleLysSerMetIleProGlu 946
 QY 1837 AGTGGCCAACTCGTGTATTTCCATAGTATACCTGCTTACCTCTCCGTCCTCAAGAA 1896
 Db 947 SerGlySerSerSer-----HisThrGlu-----LysGlu 956
 QY 1897 ATGGAAGTAGATGATGAAGATGAAGATCTGTAATG----- 1935
 Db 957 GluGluAsnGluGluLysGluLysLysProGluGlnHisLysGlnGluAspGln 976
 QY 1936 ---CTAAGAGAAACCACTTACACAAATTGAA 1965
 Db 977 GluLysArgGluLysValValAspAspMetGlu 987

RESULT 4

T42631

breast cancer tumor suppressor BRCA2 - dog (fragment)

C;Species: Canis lupus familiaris (dog)

C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 24-Nov-2003

C;Accession: T42631

R;Bignell, G.R.; Micklem, G.; Stratton, M.R.; Ashworth, A.; Wooster, R.

submitted to the EMBL Data Library, July 1996

A;Description: BRCA2 contains a novel repeat that is conserved in other mammian species.

A;Reference number: Z22184

A;Accession: T42631

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1507 <BIG>

A;Cross-references: EMBL:275664; NID:e96328; PID:e251256; PIDN:CAA99994.1

C;Genetics:

A;Gene: BRCA2

C;Superfamily: breast/ovarian cancer tumor suppressor protein, BRCA2 type

C;Keywords: tumor suppressor

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 0.000431 | Length: | 1507 |
| Score: | 166.00 | Matches: | 180 |
| Percent Similarity: | 33.96% | Conservative: | 143 |
| Best Local Similarity: | 18.93% | Mismatches: | 384 |
| Query Match: | 2.16% | Indels: | 244 |
| DB: | 2 | Gaps: | 43 |

US-09-874-162a-7 (1-4409) x T42631 (1-1507)

QY 118 TTCCCCAGCCTCGCGACCTCATCGAGCACATCGAGACCAACCATCGATACAGATCCA 177

Db 373 PheLysThrAlaSerAsnLysLysLysLysLysLysLysLysLysLysLysLysLys 392

QY 178 CGGGTTTGAAGAAACAGATTAACAGAGCCAACTAGTTGCC---CTGAGTTACATA 234

Db 393 MetLeuPheLysAspIleGluHisTyrProThrAsnLeuAlaCysLeuLysLysVal 412

QY 235 AATAGATTTCATCAGATGCTGCCCGCAGAGAGGAGTCCCTAAAGAAAGATTCAG 294

Db 413 AsnThrSerSerLeuGluSerGlnLysLysProSerLysSer-----HisAlaLeuAsp 430

QY 295 CGCAGACTCTCGTCACTCTGTCCAGCTCAGTGTCTCGAGGGAATGTGTCCACTCCCCA 354

Db 431 ProGlnSerIleAsnIleIleSerGlyPheValGln-----AsnSerThrTyrValSer 448

QY 355 CGCCACAGAGTGGAGAGCTTACTCCCGCGTAGCCACCCACCATCACCCTCTCTTCA 414

Db 449 AspSerGluSerGlyHisThrAlaProThrLeuSerLeuLysGlnAspPheAspSer 468

QY 415 TTCGCGACGACGACTCCGACAGACCAACACAGATC----- 450

Db 469 AsnArgAsnLeuThrProSerGlnLysAlaGluIleThrGluLeuSerThrIleLeuGlu 488

QY 451 -----TATAGATTCTTCGAATCTCGAATCTCATAGCATCTCATAGCACAATTTTG 495

Db 489 GluSerGlySerGlnPheGluPheThrGlnPheArgLys-----ProSerHisIle 505

QY 496 CACAGAACTTCTTACTTACATGCTCATCGAAACTCCAGAAC-----AACATCAAAAG 549

Db 506 IleGlnLysAsnProPheGluMetProGluAsnGlnLeuThrIleLeuAsnSerThrSer 525

QY 550 AAAACATTTAAAGTTGATGATATG-----TTATCAAAAGTA 585

Db 526 LysGluTyrLysAspAspAspLeuHisLeuThrAsnAlaProSerIleSerGlnVal 545

QY 586 GAGAAATGAAA-----GGAGAGCAGAGATCTCATAGCTTGTGACGT 627

Db 546 AspSerLysLysSerGluGlyIleIleGlyGlnLysPheAlaCysLeuSerArg 565

QY 628 -----CATTTGCAGCTTACGTTTACT 648

Db 566 ThrSerCysAsnArgSerAlaSerGlyTyrSerThrAspLysAsnGluValGluPheArg 585

QY 649 GGTTCCTTCCAC---AAAAATGATAGCCATCACCNAACTCAGAA----- 690

Db 586 GlyPheThrSerAlaArgGlyThrLysLeuAsnValGlySerGluAlaLeuGlnLysAla 605
QY 691 -----ATGAACAAAAATCTGTACCTCGTAACTCGAAGTC 720
Db 606 LysLysLeuPheSerAspLeuGluAsnIleAsnGluGluThrSerValGluValAsp--- 624
QY 721 CTGCTTGTAAGTGGTGGCACAACAAAGAGAGATGTAAGTTGTCCATAAGGCAAGTT 780
Db 625 -----ArgSerPheSerSerSerLysThrAsnAsp 634
QY 781 CCCACAGGTAAAGCAGTGGCTTTGATCTCTGACCTCAATCAACAAACCCGGAAT 840
Db 635 SerValSerMetIleGlnIleGluAspCysAsnAspLysAsnLeuAsnGluProAsnAsn 654
QY 841 TTCCTGCTCCCTGCGATCTCCAGTATGATGATTTGACCTAGTAAACAGCCATATGTTGAAG 900
Db 655 LysCysArgLys---IleLeuGlnAsnAsnIleGluMetThrThrAspIlePheValGlu 673
QY 901 TCTTACTCTGCTGCTATTTAGA-----GTGACTCGTCCA 933
Db 674 GluThrThrGluSerThrArgArgAsnThrGluAsnGluGluAsnGlnCysThrAspAla 693
QY 934 GGAAGAGAGAGATTTAATGGAATGATTAATGA-----GAAACCAATGAAATATT 984
Db 694 GlyArgAsnThrCysAsnSerGluSerAspGlySerAspSerSerLysAsnAspThrVal 713
QY 985 GATGTCATGAAGAG-----CTTCCAGCCAGAAAGAAACCGAAATCGTGAGGATGG 1035
Db 714 TyrIleHisGluGluAsnGlyLeuProCysIleAspGlnHisAsnIleAsp----- 731
QY 1036 GAAAGACATTTGTCGCAAAATGACAGATTTGATTTGATAAAACAGCGCTTACAGCTTTTA 1095
Db 732 -----LeuLysLeuPheSerGlnPheMetLysGluGlyAsnThrGlnIleLys 747
QY 1096 GATGGGAATATGAAGTA-----GCCATGCAGGAATCGAAGAATGTCATA 1143
Db 748 GluGlyLeuSerAspLeuThrCysLeuGluValMetLysAlaGluGluThrSerHisVal 767
QY 1144 ACCAAGAAAAGCAACATGGAGACTATTCTTGATGGGAAGAGCGCTCCCTCCATTCGAA 1203
Db 768 ThrMetSerAsnLysGlnGlnLeuThrAlaAsnThrGlyGlnAsnIleLysAspPheAsp 787
QY 1204 ACATTT-----TCTCAGGACCTACCTGTCAGTTCACTCTCTGTTGACAGGAGAGACC 1257
Db 788 ThrPheThrLeuSerPheGlnThrAlaSerArgLysAsnIleArgValSerArgGluSer 807
QY 1258 AATGATAATCTACGCGCTTATTCGCAACCTCTTGCACATAGAAAT----- 1305
Db 808 LeuAsnLysAlaArgSerLeuLeuAsnGlnLysThrThrGluGluLeuAsnAsnPhe 827
QY 1306 TCAGAGAGTCCATCAGGAAAACAAAGCTGGT---TCAGTTAAACCTTACTCAAACTATT 1362
Db 828 SerAspSerLeuAsnSerGluLeuLeuProGlyIleAspIleLysLysThrAsp----- 845
QY 1363 GCTGTTAAAGATCATCTGACTACAGATCTACAAACAGAAAAGAAAGATCTCCAAAT 1422
Db 846 ---IleSerAsnHisGluValIleGluAsnThrGluArgLysAspLysIleThrLysGlu 864
QY 1423 GAAACCCGACAAAATAAGAAATATTTTTCAGTTTCTCTATAACAACAATACAAGGCAA 1482
Db 865 SerAspLeuIleGlyThrGluAsnIleLeuLeuIleLeuGlnArgProGluSerLys 884
QY 1483 CAAACTGAAGCAGAGATGACCTGATTCCTGCTGGTGTATCTGAACTGCGGCAACTTT 1542
Db 885 IleLysLysIleLysGluSerAlaValLeuGlyPheHisThrAlaSerGlyLysLysIle 904
QY 1543 TATAGTTTACTCAACATCTTAAACTCTGCCATACAGATTT----- 1584
Db 905 GluIleThrLysGluSerLeuAspLysValLysAsnLeuPheGluGluLysGluGlnAsp 924
QY 1585 -----ATCTCAACTATGTTTATCATCCAAAAGGTGTAGATAGATGTTTCTATC 1635
Db 925 AsnSerGluIleThrAsnPheSerHis-----ArgGlyAlaLysMetSerLysAspArg 942

QY 1636 AATCAGTGTATGATGC---TCCTATGCAGAAATCTCAGGATATTCATCGCAACCT 1692
Db 943 GluGluCysLysAspGlyArgGluLeuAlaCysGlyThrThrGluIleThrThrPro 962
QY 1693 GGATTTCCTTTTACTCCCAACGACAGTAAAGAGAACACCTATCACACAT---ATTCTT 1749
Db 963 GluThrGluGluThrHisSerSerLeuGluLysLysLysLeuValSerAsnGluIleAla 982
QY 1750 GTGTGCGGCCA-----AAACGAAACA----- 1770
Db 983 AlaLeuArgProArgLeuLeuSerAspAsnLeuTyrLysGlnThrGluAsnLeuLysIle 1002
QY 1771 -----AAACGACGATGTCTGAATTTCTTGAATCTGAAGTGGGAA 1812
Db 1003 SerAspHisAlaSerGlnLysValAspValHisGluAsnThrGluLysGluThrAlaLys 1022
QY 1813 -----GTGAACACAGAAACATATAGTAGTGGCCACCAATCGT---CTGTAT 1857
Db 1023 LysProThrMetThrThrAsnGlnSerThrTyrSerAlaIleGluAsnSerProLeuThr 1042
QY 1858 TTCATAGTAGTACCTGCTTACCTCTCCGTCACAGAAATGGAAGTAGTAGTAAGAT 1917
Db 1043 PheThrGlnAspThr-----GluGluLysPheSerValSerGluAla 1056
QY 1918 GAA-----AAGGATCCTGATCGCTAAGAGAAAACCACTTACACAAATTGAAGATT 1971
Db 1057 SerLeuPheGluAlaLysLysTrpLeuArgGlu-----GlyGluTrp 1070
QY 1972 TCTGAT-----GTTAATGAGAGAGAGAAAGAGATGATGAACTCTGG----- 2013
Db 1071 AspAspGlnSerGluArgIleAsnAlaAlaLysValAsnCysLeuLysGluThrProAsp 1090
QY 2014 -----AATCTCCATGTCATGACGATGGTGTATTGCTGCAATCAA 2055
Db 1091 AspTyrValGluAsnProSerCysGlyAsnSerSerAsnSerAlaIleThrGluAsnAsp 1110
QY 2056 ATGAATCATGCTGATGCTGTTGTAGAAAATTATGGACAGAAAATAATTAAGAAAT 2115
Db 1111 LysAsnHis-----LeuSerGluLysGlnGlySerThrTyrLeuSerAsnSer 1126
QY 2116 TTATGTCGAAAC---TTCATGCTTCTATCTAGTCAGCATGATGAC-----TTTAAT 2163
Db 1127 ThrMetSerAsnSerTyrSerTyrHisProGlyPheCysHisSerSerGluValTyrAsn 1146
QY 2164 CTTATTAGCATAATGTCA-----ATAGATAAAGCTGTACC 2199
Db 1147 LysSerGluTyrLeuSerArgSerLysIleAspAsnSerGlyIleGluProValIleLys 1166
QY 2200 AAGCTCCGTGAAATGCAG----- 2217
Db 1167 AsnIleArgGluArgLysAsnIleGlyPheSerGluIleMetSerProGlyArgGluAla 1186
QY 2218 -----CAAAAATTAGAAAGGGGAA 2238
Db 1187 AspThrAspProGlnSerValAsnGluAspIleCysValGluLysLeuAlaThrAsnSer 1206
QY 2239 TCTGCTTCCCTCCAAAC----- 2256
Db 1207 SerCysLysAsnLysAsnThrAlaIleLysValAlaIleSerAspSerAsnAsnPheAsn 1226
QY 2257 -----GAAGAAATACTGAAGAACAAAATGGGACAGCAAAATGGATTTAGTGAATAAC 2310
Db 1227 ThrIleGlnLysLeuAsnSerAspSerAsnAsnSerValProAlaTyrSerThrValAsn 1246
QY 2311 TCAAAAGAG-----AAAGCTTTGAA-----ACAGATATGTC 2343
Db 1247 SerLysArgValPheValAlaHisGlnThrLysValThrGluGlyPheThrAspAsnCys 1266
QY 2344 TCAGGGTTTCAAAACAGAGCAAAAAACAAAA 2376
Db 1267 SerMetValThrLysGlnAsnThrLysSerLys 1277

RESULT 5

S38173
 myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
 N/Alternate names: protein YKR095W; protein YKR415
 C/Species: Saccharomyces cerevisiae
 C/Date: 03-May-1994 #sequence revision 03-May-1994 #text_change 21-Jul-2000
 C/Accession: S38173; S40647; S31207
 R/Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo, submitted to the Protein Sequence Database, March 1994
 A/Reference number: S38158
 A/Accession: S38173
 A/Molecule type: DNA
 A/Residues: 1-1875 <BAL>
 A/Cross-references: EMBL:Z28320; NID:5486586; PID:5486587; MIPS:YKR095W
 R/Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Jil Yeast 9, 1349-1354, 1993
 A/Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chrom
 A/Reference number: S40644; MUID:94205265; PMID:8154186
 A/Accession: S40647
 A/Molecule type: DNA
 A/Residues: 1-1875 <BOU>
 A/Cross-references: EMBL:X73541; NID:9450550; PIDN:CAA51948.1; PID:9450554
 R/Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D. Mol. Gen. Genet. 237, 359-369, 1993
 A/Title: A new yeast gene with a myosin-like heptad repeat structure.
 A/Reference number: S31207; MUID:93247549; PMID:8483450
 A/Accession: S31207
 A/Molecule type: DNA
 A/Residues: 1-300, 'A', 302-1875 <KOE>
 A/Cross-references: EMBL:LO1992; NID:6171958; PIDN:AAA34783.1; PID:6171959
 C/Genetics:
 A/Gene: SCD:MLP1
 A/Cross-references: SGD:S0001803; MIPS:YKR095W
 A/Map position: 11R

Alignment Scores:
 Pred. No.: 0.00103 Length: 1875
 Score: 161.00 Matches: 167
 Percent Similarity: 36.63% Conservative: 129
 Best Local Similarity: 20.67% Mismatches: 297
 Query Match: 2.09% Indels: 216
 DB: 2 Gaps: 35

US-09-874-162A-7 (1-4409) x S38173 (1-1875)

| | | | |
|----|-----|--|-----|
| QY | 410 | CTTCATTCGCGAGCAGCAGCTC---CGACAGACCAACACAGATCTATAGATTCTTCGAA | 466 |
| Db | 191 | LeuAsnSerArgThrGluLeuGluArgLysThrGlnGluLeuThrLeu-LeuGlnSerAs | 210 |
| QY | 467 | CTCGGAATCTCATACACCAATATTTTTCGACAGACTCTTACTTACATGTCTCATCGA- | 525 |
| Db | 210 | nAsnAspTrpLeuGluLysGluLeuArgSerLysAsnGlu-GlnTyrLeuSerTyrArgG | 230 |
| QY | 526 | -----NACTCCAGAACAAACATCAAAAGGAAACAT | 556 |
| Db | 230 | InLysThrAspLysValIleLeuAspIleArgAsnGlnLeuAsnArgLeuArgAsnAsp | 250 |
| QY | 557 | TTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGAAGAGAGACAGCAAGATCTCATA | 616 |
| Db | 250 | heGlnMetGluArgThrAsnAsnAspValLeuLysGlnLysAsnGlnLeuSerLys | 270 |
| QY | 617 | GTTTGTGACGCTATTGAGGCTTACGTTTACTGGTTTCTTCCACAAAATGATGACCAT | 676 |
| Db | 270 | erLeuGlnGluLysLeu---LeuGluIleLysGlyLeu-----SerAspS | 284 |
| QY | 677 | CACCAAACTCAGAAATGAACAAAATCTGTGTACCTCGAAGTCTGCTGTCTGAAAGTTT | 736 |
| Db | 284 | erLeuAsnSerGluLysGlnGluPheSerAlaGluMetSerLeu----- | 298 |
| QY | 737 | GCCACAAAAGAAAGAGTGTAACTTGTTCCTCCATTAAGCAAGTCCACACAGTAAAGC | 796 |

| | | | |
|----|------|--|------|
| Db | 299 | -----LysGlnArgLeuValAspLeuLeuGluSerGlnLeuAsnAlaValLysGluG | 316 |
| QY | 797 | AGTGCTCTTTGATTCCTGCACCTCAATCAACAAACACCGGAAATTTCCGCTCCCTTGAG | 856 |
| Db | 316 | luLeuAsnSerIleArgGluLeuAsnThrAlaLys-----ValIleAlaA | 331 |
| QY | 857 | TTTCCAGTAATGAATTTGAACCTAGTAAACGCCATATGTTGAAGTCTTACTGCTGTAT | 916 |
| Db | 331 | spAspSerLysLysGlnThrProGluAsnGluAspLeuLysGluLeuGlnLeuThrL | 351 |
| QY | 917 | TTAGAGTGACTCGTCCAGGAACAGAGAGTTTAATGGAATGATTAAATCGGAAACCAATG | 976 |
| Db | 351 | ysGluLysLeuAlaGlnCysGluLysGluCysLeu---ArgLeuSerSerIleThrAspG | 370 |
| QY | 977 | AAAAATTGATTCATCAATGAAGAGCTTCACGCC- | 1008 |
| Db | 370 | luAlaAspGluAspAsnGluAsnLeuSerAlaLysSerSerSerAspPheIlePheLeuL | 390 |
| QY | 1009 | -----AGAAAGAAACGAAATCGTGAGATCGGGAAGACAT | 1045 |
| Db | 390 | ysLysGlnLeuIleLysGluArgThrLysGluHisLeuGlnAsnGlnIleGluThrP | 410 |
| QY | 1046 | TTGTTGCACAAATG- | 1090 |
| Db | 410 | heIleValGluLeuGluHisLysValProIleAsnSerPheLysGluArgThrAspM | 430 |
| QY | 1091 | TTTTAGATGGGAATATGAAGTAGCCATGCAGGAAATGGAAGATGTCCAATAGCAGA | 1150 |
| Db | 430 | etLeuGluAsnGluLeuAsnAsnAlaLeuLeuLeuGluHisThrSerAsnGluLysA | 450 |
| QY | 1151 | AAAGAGCAATCGGAGAGCTATTCCTGATGGAGAGCTGCTCCATTCGAAACATTTT | 1210 |
| Db | 450 | saLaLysValLysGluLeuAsnAlaLysAsnGlnLysLeuValGluCysGluAsnAspL | 470 |
| QY | 1211 | CTCAGGACCTACG- | 1243 |
| Db | 470 | euGlnThrLeuThrLysGlnArgLeuAspLeuLysCysArgGlnIleGlnIleThrLeuL | 490 |
| QY | 1244 | GGACAGGAGAGACCAATGATAATCTACGGCTCTCTATTGCCAAA- | 1287 |
| Db | 490 | hrAsnSerValSerAsnAsp---SerLysGlyProLeuArgLysGluGluIleGlnPheI | 509 |
| QY | 1287 | ----- | 1287 |
| Db | 509 | leGlnAsnIleMetGlnGluAspAspSerThrIleThrGluSerAspSerGlnLysValV | 529 |
| QY | 1288 | -----CCTCTGCCACTAGAAATTTCA | 1308 |
| Db | 529 | aiThrGluArgLeuValGluPheLysAsnIleIleGlnLeuGlnGluLysAsnAlaGluL | 549 |
| QY | 1309 | -----GAGAGTCTCCATCAGGAAAAACAAGC | 1333 |
| Db | 549 | eulLysValValArgAsnLeuAlaAspLysLeuGluSerLysGluLysLysSerLysG | 569 |
| QY | 1334 | CTGGTTACGTTAAA---CCTACTCAAACTATT---GCTGTTAAAGATCATTTGACTACA- | 1386 |
| Db | 569 | InSerLeuGlnLysIleGluSerGluThrValAsnGluAlaLysGluAlaIleIleThrL | 589 |
| QY | 1387 | -----GATCTACAAACAAGAAAAAGAAAG | 1410 |
| Db | 589 | eulYsSerGluLysMetAspLeuGluSerArgIleGluGluLeuGlnLysGluLeuGluG | 609 |
| QY | 1411 | -----GATACCTCCAAATGAAACCCGACAAATAAGATAATTTATCATGTTTC | 1459 |
| Db | 609 | luLeuLysThrSerValProAsnGluAspAlaSer----- | 620 |
| QY | 1460 | TCTATACACATACA---AGCCACAACTGACCAAGACATGACCTGCATTCGCCCTT | 1516 |
| Db | 621 | --TyrSerAsnValThrIleLysGlnLeuThrLysArgAspLeuGluSerGlnV | 640 |
| QY | 1517 | GGTGTACTCTGAAC-----TGCCGCAAACTTTATAGTTTACTCAAGC | 1558 |
| Db | 640 | alGlnAspLeuGlnThrArgIleSerGlnIleThrArgGluSerThrGluAsnMet-Ser | 659 |


```

Db      167  erSerAla-----|||||-----ThrSerProSerT 174
Qy      680  CAAACTCAGAAATGAAACAAATTTCTGTACCTGGAAGTCCCTCTGTGTAAGATTTGCC 739
Db      174  yrSerSerTyraSerSerSerGlnTyraSp-----|||-----LeuTyrlsPheAspV 190
Qy      740  ACAAACAAAGAAAGATGATAGTTGTCATTAAGCGCAAGTTCACACAGGTAAGAAAGCAGG 799
Db      190  alThrLysLeuLysAsn-----|||-----LeuLysGluAsnSerSerAsnLeuLeuGlnL 206
Qy      800  TGCCTTTGATTCCTGACCTCAATCAACAAACCCCGGA-----837
Db      206  euProLeuPhe-----|||-----leAsnThrThrGluAlaGluPheAlaAlaSerValGlnA 224
Qy      838  -----AATTCCTGCTCCCTGCGAGTTCCAGTTCCAGTAATGAATTTGAACCTAGTAAC- 885
Db      224  rgTyrgluLeuAsnMetLysAlaLeuAsnLeuAsnSerGluSerLeuGluAsnSerSerV 244
Qy      886  -----AGCCATATGCTGAAAGTCTTACTCGTTCGTCTATTAGAGTGACTC 928
Db      244  alGluLysSerSerAlaHisHisThrLysSerHisSerIle-----258
Qy      929  GTCCAGGAAGAGAGAGATTTAATGAATGATTAATGGAGAAACCAAT-----GAAATATTG 985
Db      259  -----ProLysHisAsnGluGluValLysThrGluThrHisGlyGluGluGlnA 275
Qy      986  ATGTCATTAAGAGAGCTCCAGCCAGAGAAAGAAACGAGATCGTGAGAGTGGGAAAAGACAT 1045
Db      275  spAlaHisAspLysLysProHisAlaSerLysAspAlaHisGluLeuLysLysLysThrG 295
Qy      1046  TTGTTGCAAAATACAGATATTGTATAAAACAGCGCTTTACAGCTTTTAGATGGGAAT 1105
Db      295  luValLysLys-----298
Qy      1106  ATGAAGTAGCCATGAGGAATGGAAGATGTCATTAAGCAAGAAAGACACATGGG 1165
Db      299  --GluAspAlaLysGlnAspArgAsnGluLysValIleGlnGluProGlnAlaThrValL 318
Qy      1166  AGACTATTCTTGATGGGAAGAGCTCGCTCCATTCGAAACATTTTCTCAGGACCTACGT 1225
Db      318  euProValValAspLysLysGlu-----ProGluGluSerValGluGluAsn-----333
Qy      1226  TCAGTTCACTTCTGTCGACAGAGACACATGATAATCTACGGCTCTCTATTGCC- 1284
Db      334  -----ThrSerLysThrSerSerProSerProSerProSerProAlaA 347
Qy      1285  -----AAACCTCTTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAACCAAGC 1333
Db      347  laLysSerTrpSerAlaIleAlaSerAspAlaIleLysSerArgGlnAlaSerAsnLysT 367
Qy      1334  CTGGTTCAAGTAAACCTACTCAAACTATGCTGTTTAAAGATCATCTGACTACAGATCTAC 1393
Db      367  hrValSerGlySerMetValThrLysThrProIleSerGlyThrThrAlaGlyValSerS 387
Qy      1394  AAACAGAAACAAAGGATCTCCAAATGAAACCGACAAAATTAAGATATTTTATC 1453
Db      387  erThrAsnMetAlaAlaThrIleGlyLysSerSerSerProLeu-----402
Qy      1454  AGTTTCTTATAACAAATACAGGCAACAACTAAGCAAGACAGATGACCTGCTGATGCC 1513
Db      403  -----LeuSerLysGlnProGlnLysLysAspLysLysLysValP 416
Qy      1514  CTGTGTGACTCTGACTCCGCCCAACTTATGATTTTACTCAAGCATCTTAACTCTGCC 1573
Db      416  roProSerThrLysGlyLeuGluProLeuGlySerIleAla-----LeuArgMetCysP 434
Qy      1574  ATAGCAGATTTATCTTCAACTATGTTTATCATCCAAAAGGTGCTAGGATAGATGTTCTA 1633
Db      434  heAspProAspPheIleSerTyValLeuArgAsnLysAspValGluAsnLysIleProV 454
Qy      1634  TCAATGAGTGTATTGATGGCTCTCTATGACGAGGAATCCTCAGGATATTTCATGCCAACCTG 1693

```

```

Db      454  alHisSerIleIle-----ProArgGlyIleIleAsnArgAlaA 467
Qy      1694  GATTTCCTTTTGTAGTCGAACGAGCAGTTAAGAGAACACCTATCACATATTCTTGTGT 1753
Db      467  snIleCysPhe-----MetSerSerValLeuGlnValLeuLeuTyrc 481
Qy      1754  CGAGGCCA-----AAACGACAAACAAAGCAAGCATG- 1782
Db      481  yslsPheProPheIleAspValIleAsnValLeuSerThrArgAsnThrAsnSerArgValG 501
Qy      1782  -----1782
Db      501  lyThrSerSerCysLysLeuLeuAspAlaCysLeuThrMetTyrlsLysGlnPheAspLysG 521
Qy      1783  -----TCTCAATTTCTTGAATCTGAAGATGGG-----GAACTAGAACACAGCAAA 1825
Db      521  luThrTyrgluLysLysPheLeuGluAsnAlaAspAlaGluLysThrThrGluSerA 541
Qy      1826  GAACATATAGTAGTGGCCACAATCGTCTGTATTTCCTAGTAGTACCTGCTTACCTCTCC 1885
Db      541  spAlaLysLysSerSerLysSerLysSerPheGlnHisCysAlaThrAlaAspAlaValL 561
Qy      1886  GTCCCAACAAGATGGAAGTAGATAGTGAAGATGAAAGATCCTGATGGCTAAGAGAAA 1945
Db      561  ysProAspGluPhe-----TyrL 567
Qy      1946  AAACCATTTACAAATTTGAAGATTTTCTGATGTTAATGAAGAGAGAGAAAGATGATGA 2005
Db      567  ysthrLeuSerThrIleProLysPheLysAspLeuGlnTrpGlyHisGlnGluAspAlaG 587
Qy      2006  AACTCTGGATCTCCATCTCATGAAGCAT-----GGGT 2038
Db      587  luGluPheLeuThrHisLeuLeuAspGlnLeuHisGluLeuLeuSerAlaIleAspG 607
Qy      2039  TTATTCCTCACAATCAATGAATCATGCTGTATGCTGTTGTGAGAAATATTAGCAGAGA 2098
Db      607  lyleuThrAspAsnGlu-----IleGlnAsnMetLeuGlnS 619
Qy      2099  AATAATTAAGAGAAATTTATGTCGAAACTTTCATCTCTCATCTAGTCAGCATGCAT- 2154
Db      619  erIleAsnAspGluGlnLeu-----LysValPhePheIleArgAsnLeuSerArgTyrglyL 638
Qy      2155  -----CACTTTAATCTTATTAGCATAATGTCATAGATAAAGCTGTTTACCAGCTCCGTG 2209
Db      638  ysAlaGluPhe-----IleLysAsnAlaSerProArgLeuLysG 651
Qy      2210  AAATGACAGAAAATTTAGAAAAGGGGGAATCTGCTTCCCTGCCAACGAGAAATAACTG 2269
Db      651  luLeuIleGluLysTy-----GlyValIleAsnAspAspSerThrG 665
Qy      2270  AGAACAAAATGGACAGACAAATGGATTTAGTGAATTAACCTCAAAAGAGAAAGCTTTGG 2329
Db      665  luGlu-----AsnGlyTrpHisGlu-----671
Qy      2330  AAACAGATAGTGTCTCAGGGGTTTCAAAACAGACAGCAAAACAAAAA 2376
Db      672  -----ValSerGlySerSerLysArgGlyLysLysThrLys 683

```

RESULT 7

T15496
 hypothetical protein C14F5.3 - *Caenorhabditis elegans*
 C;Species: *Caenorhabditis elegans*
 C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 R;Accession: T15496
 R;Minx: P.

Submitted to the EMBL Data Library, June 1995
 A;Description: The sequence of *C. elegans* cosmid C14F5.
 A;Reference number: Z18361
 A;Accession: T15496
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1263 <MIN>
 A;Cross-references: EMBL:U29082; NID:g861384; PID:g861386; PIDN:AAA68402.1; CESP:C14F5.3

1765 QY -----CGAACAAAGCA-----AG 1778

656 Db lHisSerIleLysProIleAspLeuGluAsnThrLysSerArgGlnGlnAlaIleLysAs 676

1779 QY CATGTCGTGAATTTCTGAATCTCAAGATCGGGAGTAGAACAAGCAAGAAACATATAGTAG 1838

676 Db pLeuAsnGluPheLeuLysAsnProAsnAspAlaGlnAlaSerLysThrLeuAlaG1 696

1839 QY TGCGCCAACT-----CGTCTGTAATTCATAGTGTACTCTGCT 1877

696 Db nAlaAsnLysIleGlnHisLeuGluAspLeuLysSerLysValHisSer----- 712

1878 QY ACCTCTCCGTCCTCCACAGAAATGGAAGTAGTAGTAGAGATGAAAGCATCTCTGAATGGCT 1937

713 Db -----IleLysPro-----IleAspLeuGluAsnThrLysSer----- 723

1938 QY AAGAGAAAAACCATTTACACAAATTCGAAGTTTTCTGATGTTAATCAAGAGAGAAAGA 1997

724 Db ArgGlnGlnAlaIleLysAspLeuAsnGluPheLeuLysAsnProAsnAspAlaG1 743

1998 QY AGTGATGAAACTCTGGAATCTCCATGTCATGAAGCATGGGTATTATCTGCACAAATCAAT 2057

743 Db nAlaSerLysThr-----LeuAlaGlnAlaTyrg1 753

2058 QY GAATCATGCTGTATGCTGTTTGTAGAAAATTTATGGACAGAAAATTAATTAAGAAATTT 2117

753 Db uAsnAsnGlyAspLeuLeuLysAlaGluAsnAlaTyrg1uLysIleIleLysLeuThrAs 773

2118 QY ATGTGCA-----AACTTCATGCTT-----CA 2138

773 Db nThrGlnGluAspHisTyLysLeuGlyIleIleArgPheLysLeuLysLysTyrg1uHu1 793

2139 QY TCTAGTCAGATGCATGCATTTAATCTT----- 2166

793 Db sSerIleGluSerPheAspGlnThrIleLysLeuAspProLysHisLysLysAlaLeuHu1 813

2167 QY -----ATTAGCATATGTCATAGATAAGCTGTTACCAAGCTCCGTGAAATGCA 2216

813 Db sAsnLysGlyIleAlaLeuMetMetLeuAsnLysAsnLysLysAlaIleGluSerPheG1 833

2217 QY GCAG-----AAATTAGAAAAAGGGGAATCTGCTTCCCTCGCAACAGAGAAATAACTGA 2270

833 Db uLysAlaIleGlnIleAspLysAsnTyrg1yThrAlaTyrg1nLysGlyIleAlaG1 853

2271 QY AGAACAAAATGGACAGCAATGATTTAGTGAATTAACATAAAGAGAAAGCTTTGGA 2330

853 Db uGluLysAsnGly----- 857

2331 QY AACAGATAGTGTCTCAGGGGTTTCAAAACAGAGCAAAAAACAAAAAACTCTGAAAAAGCTCT 2390

858 Db Asp-MetGlnGlnAlaPhe-----AlaSerPheLysAsnAlaTyrg1AsnLeu- 872

2391 QY AACCCCATGTTATGGACAAACATGAAATTCATT----- 2425

873 Db -----AspLysAsnProAsnTyrg1AlaLeuLysAlaGlyIleValSerAsnA 888

2426 QY TTAGGGAAT---TCATCTCTAGAAATATGTTTCTGTTTTTAAT 2467

888 Db snLeuGlyAsnPheLysGlnSerGluTyrg1LeuAsnPhePheAsn 903

RESULT 9

T39068

coiled coil protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39068

R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Connor, R.E.; Wood, R.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z21825

A:Accession: T39068

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

| | | | |
|----|------|--|------|
| Db | 666 | rValLeuGlnArgGlnLeuThrAspValLysAspGluLeuAspValSerGluGluGluAr | 686 |
| Qy | 1745 | -----TTCTTGTGTCCGCCCAAAACGACAAACAAAGCATGTCTGTAATTTCTGAA | 1797 |
| Db | 686 | gGluGluAlaIleValAlaGlyGln--LysLeuSerAlaSerPhe--GluLeuMetSer | 704 |
| Qy | 1798 | TCTCAAGATGGGAAGTAGACACGACCAAGAACATATAGTAGTGCCCAATCTCTGTAT | 1857 |
| Db | 705 | AsnGlnLysGlnAlaLeuGluLeuLys-----TyrSerLeuLysAsnGluLeu--- | 721 |
| Qy | 1858 | TTCCATAGTGATACCTGTTACCTCTCCGTCCTCCACAGAAATGGAAGTAGATAGAGAT | 1917 |
| Db | 722 | IleAsnAlaGlnAsnLeuLeuAspArgGluGluGluLeuSerGluLeuSerLysLys | 741 |
| Qy | 1918 | GAAGAAGATCTGTAATGGCTAAGAGAAAAACCATTCACAAATTGAAGAGTTCCTGAT | 1977 |
| Db | 742 | LeuPheGluGluArgLysIleArgSerGlySerAsnAspIleGluLysAsnLysGlu | 761 |
| Qy | 1978 | GTTAATCAAGGAGAGAGAAGTGAATG--AAACTCTGGAATCTC--CATGTCATCAAG | 2031 |
| Db | 762 | IleAsnValLeuAsnSerGluLeuAlaAspLysLeuAlaGlnIleArgHisLeuGluSer | 781 |
| Qy | 2032 | CATGGGTTTATTGCTGACAAATCAAAATGAATCATGCTGTATGCTGTTGTAGAA--- | 2088 |
| Db | 782 | AspLysMetGluLeuAspLysLeuValHisIleAsnArgGlyIleGluGluAlaAsn | 801 |
| Qy | 2089 | TATGCACAGAAATAATTAGAAGAAATTATGTGCGAAACTTCATGCTTCATCTAGTCAGC | 2148 |
| Db | 802 | IleGluGluAsnAlaValLysLysArgLeuCys-----LeuLeuMet | 815 |
| Qy | 2149 | ATGCATGACTTTAATCTATTAGCATATATGTCATA-- | 2184 |
| Db | 816 | GlyCysAspTyrSerSerValSerIleLeuGlnIleValSerGlnIleGluHisPheVal | 835 |
| Qy | 2185 | GATAAAGCTGTATCCAGCTCCGTGAAATCAGCAAAATAATAGAA----- | 2229 |
| Db | 836 | AsnGlnGlnIleGlnThrIleArgSerLeuLysGlnGluLeuArgHisAspPheValGln | 855 |
| Qy | 2229 | ----- | 2229 |
| Db | 856 | PheSerGlyLysLysGluGlnGluLeuSerArgSerPheGluLysPheGlyLeuGlyThr | 875 |
| Qy | 2230 | -----AAGGGGGAATCTGCTTCCTCCCTGCAAAACGAGAAATACTGAAGAACAAAT--- | 2280 |
| Db | 876 | GluThrLysHisAspIleLeuAlaGlnArgAsnArgAsnValSerGluLysMetAsnAsp | 895 |
| Qy | 2281 | -----GGGACAGCAATGATTT----- | 2298 |
| Db | 896 | LeuGluAsnAlaAlaGlnLysPhePheSerSerProAspArgLysAsnGlyTyrLeuTyr | 915 |
| Qy | 2299 | -----AGTGAATTAACCTCAAAAGAGAAAGCTTTGAA----- | 2331 |
| Db | 916 | ProSerGluHisThrSerLysIleGluTyrLeuGlnLysThrIleGluAspLeuLysLeu | 935 |
| Qy | 2332 | -----ACAGATAGTGCTCAGCGGGTTTCA----- | 2355 |
| Db | 936 | AlaLeuGlnAspGluLeuLysAsnArgAsnLeuLeuMetAspAspIleSerSerTyrAsn | 955 |
| Qy | 2356 | AAACAGACAGAAAAA | 2370 |
| Db | 956 | LysGlnThrThrLys | 960 |

RESULT 10
S69838 TyB protein - yeast (*Saccharomyces cerevisiae*) retrotransposon Ty1.GR1
N:Alternate names: protein G4054
C:Species: *Saccharomyces cerevisiae*
C/Date: 09-Mar-1996 #sequence_revision 06-Sep-1996 #text_change 20-Jun-2000
C/Accession: S69838
F:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64071

A:Accession: S69838
A:Molecule type: DNA
A:Residues: 1-1755 <R1>
A:Cross-references: EMBL:Z72812; NID:G1325961; PDB:CAA37012.1; PID:G1325964
A:Note: Biosynthesis of this protein involves a +1 frameshift in the codon for residue 4
C:Genetics:
C:Map position: 7R
A:Mobile element: retrotransposon Ty1.GR1
C:Superfamily: TyB protein

| Alignment Scores: | 0.00239 | Length: | 1755 |
|--|---------|---------------|------|
| Pred. No.: | 156.00 | Matches: | 184 |
| Percent Similarity: | 35.44% | Conservative: | 124 |
| Best Local Similarity: | 21.17% | Mismatches: | 283 |
| Query Match: | 2.03% | Indels: | 278 |
| DB: | 2 | Gaps: | 48 |
| US-09-874-162A-7 (1-4409) x S69838 (1-1755) | | | |
| Qy 163 ATCGATACAGATCCACGGGTTTAA-----CAAAAACAAGAAATTACAGCAGCCCAACC 213 | | | |
| Db 881 ValAspThrThrAsnTyrValIleLeuGlnGlyLysGluSerArgLeuAspGlnPheAsn 900 | | | |
| Qy 214 TATCTTGCCTCGAGTTAC-----ATAAATAGATTTCATGACAGAT----- 252 | | | |
| Db 901 TyrAspAlaLeuThrPheAspGluAspLeuAsnArgLeuThrAlaSerTyrHisSerPhe 920 | | | |
| Qy 253 ---GCTCCCGCCGAGCAGGAGTCCCTAAAGAAGAAGATT----- 291 | | | |
| Db 921 IleAlaSerAsnGluIleGlnGluSerAsnAspLeuAsnIleGluSerAspHisAspPhe 940 | | | |
| Qy 292 -----CAGCCGAAGCTCTCGTGAGTCTGTCCAGTCA 324 | | | |
| Db 941 GlnSerAspIleGluLeuHisProGluGlnProArg-----AsnValLeuSerLysAla 958 | | | |
| Qy 325 GTGTCCTCGAGGAGTGTCCATCCGCCCA-----CGCCAC 360 | | | |
| Db 959 ValSerProThrAsp---SerThrProProSerThrHisThrGluAspSerLysArgVal 977 | | | |
| Qy 361 AGCAGTGGAGCCCTTACTCCCCC-----GTGACCCCAACCATCACCCTCC----- 408 | | | |
| Db 978 SerLysThrAsnIleArgAlaProArgGluValAspProAsnIleSerGluSerAsnIle 997 | | | |
| Qy 409 ----TCTTCATTCCGACAGCAGTCCGACAGAGCCAAACAGATCTATAGATTTCTT 462 | | | |
| Db 998 LeuProSerLysLysArgSerSerThrPro-----GlnIleSerAsnIleGlu 1013 | | | |
| Qy 463 CGAACTCGGAAT-----CTCATAGCACCAATATTTTGT 495 | | | |
| Db 1014 SerThrGlySerGlyGlyMetHisLysLysLeuAsnValProLeuLeuAlaProMetSerGln 1033 | | | |
| Qy 496 CACAGAATCTTACTTACATGTCTCATCGAAACTCCAGAACAAACATCAAAAGGAAACA 555 | | | |
| Db 1034 SerAsnThr-----HisGluSerSerHisAlaSer---LysSerLysAsp 1047 | | | |
| Qy 556 TTTTAAAGTTCGATATGTGTTATCAAAAGTAGAGAAAATGAAGAGACGACGAATCTCAT 615 | | | |
| Db 1048 PheArgHisSerAspSerTyrSer-----GluAsnGlu 1058 | | | |
| Qy 616 AGCTTGTCAGCTCATTTGCGAGCTTACGTTTACTGTGTTTCTTCCACAAAATGAAGCCA 675 | | | |
| Db 1059 ThrAsnHisThrAsnValProIleSerSerThrGly-----GlyThrAsnAsnLysThr 1076 | | | |
| Qy 676 TCACCAAAATCAGAAAATGAACAAAATTCGTTTACCCTGGAGGTCCTGCTGTGAAAGTT 735 | | | |
| Db 1077 ValProGlnIleSerAspGlnGluThrGlu-----LysArgIle 1089 | | | |
| Qy 736 TGCACCAAAAAGAAAG---GATGTAAGT----- 762 | | | |
| Db 1090 IleHisArgSerProSerIleAspAlaSerProGluAsnAsnSerSerHisAsnIle 1109 | | | |
| Qy 763 TGTCCATAAGCGCAAGTTCACACAGGTAAAGACAG----- 798 | | | |

```

Db      1110 ValProIleLysThrProThrThrValSerGluGlnAsnThrGluGluSerIleAala 1129
QY      799 ---GTGCTTTGATTCCTGACCTCAATCAACAAACCCGGAATTTCCCGTCCCTTGCA 855
Db      1130 AspLeuProLeu---ProAspLeuProProGluSerProThrGluProAsp----- 1146
QY      856 GTTCCAGTAATGAATTTGAACCTAGTAACAGCCATATGGTGAAGTCTTACTCGTCTTA 915
Db      1147 ---ProPheLysGluLeuProProIleAsnSerArgGlnThrAsnSer----- 1161
QY      916 TTTAGAGTGACTCGTCGACGAGAGAGAGAGAGTTAATGGAATGATTAATGGAAACCAAT 975
Db      1162 -----SerLeuGlyGlyIle-----GlyAspSerAsn 1170
QY      976 GAAATATTGTGATCAATCAAGAGCTTCCAGCCAGAGAAACGAAATCGTGAAGATGGG 1035
Db      1171 AlaTyThrThrIleAsnSer-----LysLysArgSerLeuGluAspAsn 1185
QY      1036 GAAAGACATT---GTTGCACAATGACAGATTTTGTATGAAACAGCGCGCTTACAGCTT 1092
Db      1186 GluThrGluIleLysValSerArgAspThrTrpAsnThrLysAsnMetArg----- 1202
QY      1093 TTAGATGGGGAATATGAAGTAGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1152
Db      1203 -----SerLeuGluProProArgSerLysLys 1211
QY      1153 AGA-----GCAACATGGGAGACTATTTGTGATGGAGAGAGGCTGCCTCCATTC 1200
Db      1212 ArgIleHisLeuIleAlaAlaValIleVal-----LysSerIleLysProIle 1228
QY      1201 GAACATTTCTCAGGACCTACGTTGCAGTTC-----ACTCTTCGTTGGACAGGAGAG 1254
Db      1229 ArgThr-----ThrLeuArgTyrAspGluAlaIleThrTyrAsnLysAsp 1243
QY      1255 ACCAATGATAAATCTACGGCTCTTATTGCCAAACCTCTTGCCACTAGAAATTCAGAGAGT 1314
Db      1244 IleLysGluLysGlu-----LysTyrIleGluAla 1253
QY      1315 CTCCTACAGGAACAAACCGCTGTTTCAGTTAAACCTACTCAAACCTATGCTGTTAAAGAA 1374
Db      1254 TyrHisLysGlu-----ValAsnGlnLeuLeuLysMetLysThr 1266
QY      1375 TCATTGACTACAGATCTACAAACAAAGAAAGAGGATCTCCAAATGAAACCGACAA 1434
Db      1267 TrpAspThrAspGluTyrTyrAspArgLysGluIleAsp-----Pro 1280
QY      1435 AAATTAAGAATATTATTCATGTTTCTCTATTAACAACATACAGGCAACAAACCTGAAGCA 1494
Db      1281 LysArgValIleAsnSerMetPheIlePheAsn-----LysLys 1293
QY      1495 AGAGATGACCTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1554
Db      1294 ArgAspGlyThrHis----- 1298
QY      1555 AAGCATCTTAAACTCTGCCATPAGCAGATTTATCTTCAACTATGTTTATCTCCAAAGGT 1614
Db      1299 -----LysAlaArgPheVal----- 1303
QY      1615 GCTAGGATAGATGTTTCTATCAATGAGTGTTATGATGCTTCCTATGCTCAGGAAATCCTCAG 1674
Db      1304 AlaArgGlyAspIleGlnHisProAspThrTyrAspSerGlyMetGlnSerAsnThrVal 1323
QY      1675 GATATTCTCCCAACCTCGGATTTGCTTTTAGTCGCAACGACCGACAG-----TT 1722
Db      1324 HisHisTyr-AlaLeuMetThrSerLeuSerLeuAlaLeuAspAsnAsnTyrTyrIleTh 1343
QY      1723 AAGAGAACACCTATCACACATTTCTGTGTGTGAGGCGCAACGAAACAAAGCAAGCATG 1782
Db      1343 rGlnLeuAspIleSerAlaTyrLeuTyrAlaAspIleLysGlu----- 1359
QY      1783 TCTGAATTTCTTGAATCTGAAGATGGGGAAGTAGAACAGCAAGAAACATATAGTAGTGGC 1842

```

```

Db      1359 ----- 1359
QY      1843 CACAATCGTCTGTTATTTCCATAGTATGATCTCTTACCTCTCGCTCCACAGAAATGGAA 1902
Db      1360 -----LeuTyrIleArgProProHisLeuGlyMetAsnAspLys----- 1373
QY      1903 GTAGATAGTGAAGATGAAGAGTCTGAATGGTGAAGAGAAACCAACCATACAC----- 1957
Db      1374 -----LeuIleArgLeuLysSerLeuTyrGlnLysGlnSerGlyAlaAsnTrpTy 1392
QY      1958 -----AAATTGAAGAGTTTCTGATGTTAATGAAGAGAGAGAAAGATGATGAACCTCGG 2013
Db      1392 rGluThrIleLysSerTyrLeu--IleGlnGlnCysGlyMetGluGluValArgGlyTyr 1411
QY      2014 AATCTCCATGTCATGAAGCATGGTTTATTTGCTGCAATCAATGAATCATGCCCTGTATG 2073
Db      1412 SerCys-----ValPheLysAsnSerGlnValThrIleCys--- 1423
QY      2074 CTGTTTGTAGAAAATTTATGGACAGAAAATAATTAAGAGAAATTTATGCGAACTTCATG 2133
Db      1424 LeuPheValAspAsp-----MetValLeuPheSerLysAsnLeuAsnSerAsnLysArg 1441
QY      2134 CTT-----CATCTAGTCAGCATGATGATGATCTTAAATCTTATTAGCATATGTCATAGAT 2187
Db      1442 IleIleAspLysLeuLysMetGlnTyrAspThrLysIleIleAsnLeuGlyGluSerAsp 1461
QY      2188 AAAGCTGTTTACC---AAGCTCGTGAATGCGCAAAAATTAAGAAAAGGGGAATCTGCT 2244
Db      1462 GluGluIleGlnTyrAspIleLeuGlyLeuGluIleLysTyrGlnArgGlyLysTyrMet 1481
QY      2245 TCCCTCGCAACAGCAAAATACTCAAGAACCAAAATGGACAGCAAAATGGATTTAGTGAA 2304
Db      1482 LysLeuGlyMetGluAsnSerLeuThrGluLysIleProLysLeuAsn-----ValPro 1499
QY      2305 ATTAATCAAAAGAGAAAGCTTTGAAAACAGATAGTGTCTCAGGGGTTTCAAAACAGAGC 2364
Db      1500 LeuAsnProLysGlyArgLysLeuSerAlaProGlyGlnProGlyLeuTyrIleAspGln 1519
QY      2365 AAAAACAACAAACTCTGAA-----AAGCTCTAACCCCATGTTATGCGACAAACA 2412
Db      1520 GlnGluLeuGluLeu-LeuGluAspAspTyrLysMetLysValHisGluMetGlnLys-- 1538
QY      2413 CTGAATTTACATTTTAGGGAATTCCTCTCAAGAAATTTATGTTTGTTTTAAATCATAT 2472
Db      1539 -----LeuIleGlyLeuAlaSerTyrValGlyTyrLysPheArgPheAspLeuLe 1555
QY      2473 GTTCCAAACAGGCACTGTT 2491
Db      1555 uTyTyrIleAsnThrLeu 1561

RESULT 11
S57045
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR
N;Alternate names: protein JUS60; protein YJ027W
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-1999
C;Accession: S57045; S60511
R;Zagulska, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herber
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57040
A;Accession: S57045
A;Molecule type: DNA
A;Residues: 1-1755 <HUA>
A;Cross-references: EMBL:Z49526
A;Note: Biosynthesis of this protein involves a +1 frameshift in the codon for residue 4;
R;Zagulska, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytka, J.; Sulicka, J.; Herber
Yeast 11, 1179-1186, 1995
A;Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading fra
A;Reference number: S60503; MUID:96109930; PMID:8619316
A;Accession: S60511
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 'TISTTFT', 435-1755 <ZAG>

```

A;Cross-references: EMBL-X87297
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
A;Note: the difference at the amino end is due to translation from an incorrect reading
C;Generics:
A;Map position: 10R
A;Mobile element: retrotransposon Ty1-JR
C;Superfamily: TyB protein

| | |
|------------------------|--------|
| Alignment Scores: | |
| Pred. No.: | 0.0026 |
| Score: | 159.50 |
| Percent Similarity: | 36.70% |
| Best Local Similarity: | 22.62% |
| Query Match: | 2.02% |
| DB: | 2 |
| | 44 |
| | 213 |
| Indels: | 2.02% |
| Mismatches: | 232 |
| Conservative: | 99 |
| Matches: | 159 |
| Length: | 1755 |

US-09-874-162A-7 (1-4409) x S57045 (1-1755)

Qy 163 ATCGATACAGNATCCACGGGTTTTA-----GAAAAACAAGAATTACACAGCCCAACC 213
Db 881 ValAspThrThrAsnTyrValIleLeuGlnGlyLysGluSerArgLeuAspGlnPheAsn 900
Qy 214 TATGTTCCCTCGAGTTAC-----ATAAATAGATTCAATCAGAGAT----- 252
Db 901 TyrAspAlaLeuThrPheAspGluAspLeuAsnArgLeuThrAlaSerTyrGlnSerPhe 920
Qy 253 ---GCTCCCGCCGAGAGCAGGATCCCTAAAGAGAAGATT----- 291
Db 921 IleAlaSerAsnGluIleGlnGluSerAsnAspLeuAsnIleGluSerAspHisAspPhe 940
Qy 292 -----CAGCCGAAGCTCTCGCTCACTCTCCAGTCTCTCCAGTCC 324
Db 941 GlnSerAspIleGluLeuHisProGluGlnProArg-----AsnValLeuSerLysAla 958
Qy 325 GTGTCTCGAGGGAATGTCCACTCCCCCA-----CGCCAC 360
Db 959 ValSerProThrAsp---SerThrProProSerThrHisThrGluAspSerLysArgVal 977
Qy 361 AGCAGTGGAAAGCCTTACTCCCCCC-----GTGACCCCACTCACCCTCC----- 408
Db 978 SerLysThrAsnIleArgAlaProArgGluValAspProAsnIleSerGluSerAsnIle 997
Qy 409 -----TCTTCATTCCGACAGACACGCCAGACGCCAACACAGATCTATAGATTCTT 462
Db 998 LeuProSerLysLysArgSerSerPro-----GlnIleSerAsnIleGlu 1013
Qy 463 CGAACTCCGAAT-----CTCATGACCAATATTTTGT 495
Db 1014 SerThrGlySerGlyGlyMethHisLysLeuAsnValProLeuLeuAlaProMetSerGln 1033
Qy 496 CACAGACTCTTACTTACATCTCTCATCGAACTCCAGACACACATCAAAAGAAACA 555
Db 1034 SerAsnThr-----HisGluSerHisAlaSer---LysSerLysAsp 1047
Qy 556 TTTAAAGTTGATGATATGTTATCAAAAGTAGAGAAAATGAAGGAGACAGAATCTCAT 615
Db 1048 PheArgHisSerAspSerTyrSer-----GluAsnGlu 1058
Qy 616 AGCTTGTCAGCTCATTTGCAGCTTACGTTTACTCGTTTCTTCCAAAAAATGATAAGCCA 675
Db 1059 ThrAsnHisThrAsnValProIleSerSerThrGly-----GlyThrAsnAsnLysThr 1076
Qy 676 TCACAAACTCAGAAAATGAACAAAATTCGTTTACCCTGGAAGTCCTGCTTGTAAGT 735
Db 1077 ValProGlnIleSerAspGlnThrGlu-----LysArgile 1089
Qy 736 TGCCACAAAAAACAAG--GATGTAAGT----- 762
Db 1090 IleHisArgSerProSerIleAspAlaSerProProGluAsnAsnSerSerHisAsnIle 1109
Qy 763 TGTCCAATAAGCGAAGTCCACAGGTAAAGCAG----- 798
Db 1110 ValProIleLysThrProThrThrValSerGluGlnAsnThrGluGluSerIleAla 1129

Development 105, 723-731, 1989
 A>Title: Cloning and characterization of a myoblast cell surface antigen defined by 24.1
 A;Reference number: A43557; MUID:90091812; PMID:2480877
 A;Accession: A43557
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 312-707 <GOW>
 A;Cross-references: GB:X16850; NID:G23711; PIDN:CAA34747.1; PID:G23712
 C;Genetics:
 A;Gene: GDB:SFPQ; PSF
 A;Cross-references: GDB:138275
 A;Map position: 4q-4q
 C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein
 C;Keywords: alternative splicing; pre-mRNA splicing; surface antigen
 F;298-359/Domain: ribonucleoprotein repeat homology <RRM1>
 F;372-438/Domain: ribonucleoprotein repeat homology <RRM2>

Alignment Scores:
 Pred. No.: 0.0035 Length: 707
 Score: 153.00 Matches: 120
 Percent Similarity: 29.95% Conservative: 60
 Best Local Similarity: 19.97% Mismatches: 207
 Query Match: 1.99% Indels: 214
 DB: 2 Gaps: 22

US-09-874-162A-7 (1-4409) x A46302 (1-707)

| | | | |
|----|-----|--|-----|
| QY | 2 | CCGCCCGCGCTCGAGAGCCGACACAGGGGGCTCGATGTAGCACCACGAGCA | 61 |
| Db | 35 | ProProGlyWetGlyLeuAsnGlnAsnArgGlyProMetGlyProGlyProGlyGlnSer | 54 |
| QY | 62 | TGCGCGCGCTCTCTCTCCATACCTGCGGATTCGGGGTGGCGGACTCCACTCC | 121 |
| Db | 55 | GlyProLysProProLysProProLysProProLysGlnGlnGlnProProPr | 74 |
| QY | 122 | CCACCTGCGCGACTCTCGACATCGAGGACCAACACATCATCATCATCATCGGG | 181 |
| Db | 74 | oGlnGlnProProGlnGlnPro-----ProHisGln----- | 86 |
| QY | 182 | TTTTAGAAAACAGAAATACAGAGCAACCTATGTTGCCGTGATACATAATAGAT | 241 |
| Db | 87 | -----ProProHisProGlnPr | 93 |
| QY | 242 | TCATGACAGATGCTCCCGCGAGAGGAGTCCCTAAAGAAAGATTCAGCGGAGC | 301 |
| Db | 93 | oHisGlnGlnGlnProPro-----ProProGlnAspSerSerLysPr | 109 |
| QY | 302 | TCTCGCTGACTCTGTCAGCTCAGTGCTCGAGGAATGTCCTCCACTCCCCCAGCCACA | 361 |
| Db | 109 | oVal-ValAlaGlnGlyProGlyProAlaProGlyValGlySerAlaProProAlaSerS | 129 |
| QY | 362 | GCAGTGGAGCTTACTCCCGCGTGACCCACCCATCACC-----CCCTCTCTTCAT | 415 |
| Db | 129 | erSer-----AlaProProAlaThrProProoThrSerGlyAlaProProGlySerG | 146 |
| QY | 416 | TCCGAGCAGCACTCCGACAGAGCCACACAGATCTATAGATTTCTTCGAACTCGGAATC | 475 |
| Db | 146 | LyProGlyProThrProThrProProProAlaVal----- | 157 |
| QY | 476 | TCATAGCACCATTATTTTGCACAGACTTTACTACATGCTCTCATCTCCGAACTCCAGAA | 535 |
| Db | 157 | ----- | 157 |
| QY | 536 | CAAAATCAAAAGGAAACATTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGA | 595 |
| Db | 157 | ----- | 157 |
| QY | 596 | AAGGAGCAAGATCTCATAGCTGTGAGCTCATTTGCAGCTTACGTTTACTGTTCT | 655 |
| Db | 158 | -----ThrSerAlap | 161 |
| QY | 656 | TCCACAAAATGATAGCCATCACCACCACTCAGAAAATGACAAAATCTGTACCCCTGG | 715 |

| | | | |
|----|------|--|------|
| Db | 161 | roProGlyAlaProProProProThrProProSerSerGly----- | 173 |
| QY | 716 | AAGTCCTGCTGTGAAAGTTTGCACAAAAAGAGGATGTAGTTGTCCAATAAGGC | 775 |
| Db | 174 | -----ValProThrThrProGlnAlaGlyGlyProProProProAlaA | 190 |
| QY | 776 | AAGTTTCCC-----ACAGGTAAAGCAGGTGCT----- | 804 |
| Db | 190 | laValProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyGlyG | 210 |
| QY | 805 | --TTGATTCCTGCTCAATCAACAAACCCGAAATTTCCGTCCTTCGCAATTTCCA | 862 |
| Db | 210 | lyLysMetPro-----GlyGlyProLysProGlyGlyProGlyLeuSerThrProG | 228 |
| QY | 863 | GTAATGATTTGAACCTAGTAACAGCCATATGTTGAAGTCTTACTCGTCTTATTAGAG | 922 |
| Db | 228 | lyGlyHisProLysProProHisArgGlyGlyGlyGlu----- | 240 |
| QY | 923 | TGACTCGTCCAGGAAGAGAGATTT----- | 948 |
| Db | 241 | --ProArgGlyGlyArgGlnHisHisProProThrHisGlnGlnHisHisGlnGlyPro | 260 |
| QY | 949 | --AATGGAATGATTATGAGAAACCAATGAAATATTGATCAATGAAGAGCTCCAG | 1006 |
| Db | 260 | roProGlyGlyProGlyGlyArgSerGluGluLysileSerAspSerGluGlyPheLysA | 280 |
| QY | 1007 | CCAGAAAAACCAATCGTGAAGTGGGAAAGACATTTGTCACAAATGACAGTAT | 1066 |
| Db | 280 | laasnLeuSerLeuLeuArgArgProGlyGluLysThrThrThrGlnArgCysArgLeuP | 300 |
| QY | 1067 | TTGATAAAACACGGCGTTTACAGCTTTTATGATGGGAATAT----- | 1107 |
| Db | 300 | heValGlyAsnLeuProAlaAspIleThrGluAspGluPheLysArgLeuPheAlaLysT | 320 |
| QY | 1107 | ----- | 1107 |
| Db | 320 | yrGlyGluProGlyGluValPheIleAsnLysGlyLysGlyPheGlyPheIleLysLeuG | 340 |
| QY | 1108 | -----GAATAGCCATGAGGAAATGGAAGAAATGCAATTAAGCAAGA | 1150 |
| Db | 340 | luSerArgAlaLeuAlaGluIleAlaLysAlaGluLeuAspAspThrProMetArgGlyA | 360 |
| QY | 1151 | AAAGA-----GCACATGGAGACTATTTGATGGGAGAGGCTGCTC | 1195 |
| Db | 360 | rgGlnLeuArgValArgPheAlaThrHisAlaAlaAlaLeuSerValArgAsnLeuSerP | 380 |
| QY | 1196 | CATTC-----GAAACATTTTCTCAG--GGACCTACCTGCTCAGT | 1231 |
| Db | 380 | roThrValSerAsnGluLeuGluGluAlaPheSerGlnPheGlyProIleGluArgA | 400 |
| QY | 1232 | TCACTCTT-----CGTGGACAGGAGAG--ACCAATGATAAATCTA | 1270 |
| Db | 400 | laValValIleValAspAspArgGlyArgSerThrGlyLysGlyIleValGluPheAlaS | 420 |
| QY | 1271 | CGGCTCTTATGCAAACTCTTGCCTAGAAATTCAGAGAGT-----CTCATC | 1321 |
| Db | 420 | erLysProAlaAlaArgLysAlaPheGluArgCysSerGluGlyValPheLeuLeuThrT | 440 |
| QY | 1322 | AGAAACAGCTGCTTCAATTAACCTACTTCAAACTATGCTGTTAAAGATCATTTGA | 1381 |
| Db | 440 | hrThrProArgProValIleValGluProLeuGluGlnLeuAspAspGluAspGlyLeuP | 460 |
| QY | 1382 | CTACAGATCTACAACAAGA-----AAAGAAAGGATATCTCAATGAAA | 1426 |
| Db | 460 | roGluLysLeuAlaGlnLysAsnProMetThrGlnLysGluArgGluThrProProArgP | 480 |
| QY | 1427 | ACGCAAAAATTAAGATATTTTATCATGTTCTC-----T | 1462 |
| Db | 480 | heAlaGlnHisGlyThrPheGluThrGluThrSerGlnArgThrLysSerLeuAspGluM | 500 |
| QY | 1463 | ATAACACAAATACAGGCAACAACCTGAA-----GCAAGAGATGACCTG | 1506 |
| Db | 500 | etGluLysGlnGlnArgGluGlnValGluLysAsnMetLysAspAlaLysAspLysLeu | 519 |

RESULT 15
 S69972
 Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.N
 N:Alternate names: protein N2453
 C:Species: Saccharomyces cerevisiae
 C:Date: 09-Mar-1996 #sequence_revision 06-Sep-1996 #text_change 20-Jun-2000
 C:Accession: S69972
 R:Berger, P.; Daignon, F.; Crouzet, M.
 submitted to the Protein Sequence Database, April 1996
 A:Reference number: S62975
 A:Accession: S69972
 A:Molecule type: DNA
 A:Residues: 1-1749 <BER>
 A:Cross-references: EMBL:Z71330; NID:g1301918; PID:CAA95924.1; PID:g1301920
 A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residue 4
 C:Genetics:
 A:Map position: 14L
 A:Mobile element: retrotransposon Ty1.N
 C:Superfamily: Tyb protein

Alignment Scores:
 Pred. No.: 0.0047 Length: 1749
 Score: 152.00 Matches: 177
 Percent Similarity: 35.59% Conservative: 118
 Best Local Similarity: 21.35% Mismatches: 264
 Query Match: 1.98% Indels: 270
 DB: 2 Gaps: 48

US-09-874-162A-7 (1-4409) x S69972 (1-1749)

| | | | |
|----|------|---|------|
| QY | 163 | ATCGATACAGATCCACGGGTTTAA-----GAAAAAACAAGAAATTACAGCAGCCCAACC | 213 |
| DB | 881 | ValAspThrThrAsnTyrValIleLeuGlnGlyLysGluSerArgLeuAspGlnPheAsn | 900 |
| QY | 214 | TATGTTGCCCTCAGTTAC-----ATAAATAGATTCATGACAGAT----- | 252 |
| DB | 901 | TyrAspAlaLeuThrPheAspGluLeuAsnArgLeuThrAlaSerTyrGlnSerPhe | 920 |
| QY | 253 | ---GTCGCCCGCAGACGAGGAGTCCCTAAAGAAAGAGATTCAGCCGAGCTCTCGCTG | 309 |
| DB | 921 | IleAlaSerAsnGluIleGlnSerAspAspLeuAsnPheGlnSerAspIleGluLeu | 940 |
| QY | 310 | -----ACTCTGCCAGCTCAGTGTCTCAGGGAATGTCTCCACT | 348 |
| DB | 941 | HisProGluGlnProArgAsnValLeuSerLysAlaValSerProThrAsp---SerThr | 959 |
| QY | 349 | CCCCCA-----CGCCACGACGTGGAGCCCTACTCCCCCC | 384 |
| DB | 960 | ProProSerThrHisThrGluAspSerLysArgValSerLysThrAsnIleArgAlaPro | 979 |
| QY | 385 | -----GTGACCCCAACCCATCCACCCCTCC-----TCTTCATCCGACGAGC | 426 |
| DB | 980 | ArgGluValAspProAsnIleSerLysSerAspIleLeuProSerLysLysArgSerSer | 999 |
| QY | 427 | ACTCCGACAGACCCAAACAGATCTATAGATTCTTCGAATCCGAAT----- | 474 |
| DB | 1000 | ThrPro-----GlnIleSerAspIleGluSerThrGlySerGlyGlyMethHis | 1015 |
| QY | 475 | -----CTCATGACCAATAATTTTGGACAGAACTCTTACTTACATGTCT | 519 |
| DB | 1016 | ArgLeuAspValProLeuLeuAlaProMetSerGlnSerAsnThr----- | 1030 |
| QY | 520 | CATCGAAACTCCAGAACAAACATCAAAAGGAAAAACATTAAGTTGATGATGATATCA | 579 |
| DB | 1031 | HisGluSerSerHisThrSer---LysSerLysAspPheArgHisSerAspSerTyrSer | 1049 |
| QY | 580 | -----AAA | 582 |
| DB | 1050 | AspAsnGluThrAsnHisThrAsnValProIleSerSerThrGlyGlyThrAsnAsnLys | 1069 |
| QY | 583 | GTAGAGAAAAATGAAAGGAGACGAAGATCTCATAGCTTGTTCAGCTCATTTGCAGCTTACG | 642 |

| | | | |
|----|------|---|------|
| DB | 1070 | ThrValProGlnThrSerGluGlnGluThrGluLysArgIleIleHis----- | 1085 |
| QY | 643 | TTTACTGGTTCTTCCACAAAAATGATAGCCATCACCAAACTCAGAAAAATGACAAAAAT | 702 |
| DB | 1086 | -----ArgSerProSerIleAspThrSerSerSerGluSer | 1097 |
| QY | 703 | TCTGTTACCTCGAAGTCTCTGTGTA-----GTTTCCCAACAAAAAGAAAG | 753 |
| DB | 1098 | AsnSerLeuHisHisValProIleLysThrSerAspThrCysProLysGluAsnThr | 1117 |
| QY | 754 | GATGTAAGTTGTCATTAAGCAAGTTCACACAGGTAAAGACAGGTGCCTTTGATTCCT | 813 |
| DB | 1118 | GluGluSerIle---IleAlaAspLeuPro-----LeuPro | 1128 |
| QY | 814 | GACCTCAATCAACAAAAACCCGAAATTTCCCGTCCCTTCAGATTCCAGTAATGAATTT | 873 |
| DB | 1129 | AspLeuProGluProPro-----ThrLysLeuSerAspSerPheLysGluLeu | 1145 |
| QY | 874 | GAACCTAGTAACAGCCATATGTTGAAGTCTTACTCTGTTGCTATTTAGAGTGACTCGTCCA | 933 |
| DB | 1146 | ProIleAsnSerArgGlnThrAsnSer----- | 1155 |
| QY | 934 | GGAAGAGAGAGTTTAATGGAATGATTAAATGAGAAACCAATGAAATATTGATGTCAAT | 993 |
| DB | 1156 | -----SerLeuGlyGlyIle-----GlyAspSerAsnAlaTyrThrIleAsn | 1170 |
| QY | 994 | GAAAGAGCTTCCAGCCAGAAACCAATCOTGAGGATGGGAAAGACATTT---GTT | 1050 |
| DB | 1171 | Ser-----LysLysArgSerLeuGluAspAsnGluThrGluIleLysVal | 1185 |
| QY | 1051 | GCACAAATGACAGTATTTGATTAATAACAGCGCTTACAGCTTTAGATGGGGAATATGAA | 1110 |
| DB | 1186 | SerArgAspThrTrpAsnThrLysAsnMetArg----- | 1196 |
| QY | 1111 | GTAGCCATCGAGAAATGGAAGATGTCCAATAGCAAGAAAGAAAGA-----GCA | 1158 |
| DB | 1197 | -----SerLeuGluProProArgSerLysLysArgIleHisLeuIleAla | 1211 |
| QY | 1159 | ACATGGGAGACTATCTTGTATGGGAAGAGGCTCCATTCGAAACATTTTCTCAGGGA | 1218 |
| DB | 1212 | AlaValLysAlaVal-----LysSerIleLysProIleArgThr----- | 1224 |
| QY | 1219 | CCTAGTTCAGTTC-----ACTCTGCTGTCAGAGAGACCAATGATAAATCTACG | 1272 |
| DB | 1225 | -----ThrLeuArgTyrAspGluAlaIleThrTyrAsnLysAspIleLysGluLysGlu | 1242 |
| QY | 1273 | GCTCCTATTGCCAAACCTTTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAACAG | 1332 |
| DB | 1243 | -----LysTyrIleGluAlaTyrHisLysGlu----- | 1251 |
| QY | 1333 | CCTGTTTCAAGTTAAACCTACTCAAACTATTGCTGTTAAAGAAATCATTTGACTACAGATCTA | 1392 |
| DB | 1252 | -----ValAsnGlnLeuLeuLysMetLysThrTrpAspThrAspLysTyr | 1266 |
| QY | 1393 | CAAAACAAGAAAGAAAGGATCTCCAAATGAAAAACCGACAAAAATTAAGAAATTTAT | 1452 |
| DB | 1267 | TyrAspArgLysGluIleAsp-----ProLysArgValIleAsnSer | 1280 |
| QY | 1453 | CAGTTTCTCTATAACCAATACA-----AGCAACAACCAACCAACAGAGAT | 1500 |
| DB | 1281 | MetPheIlePheAsnArgLysArgAspGlyThrHisLysAlaArgPheValAlaArgGly | 1300 |
| QY | 1501 | GACCTCAATTCGCT-----TGGTGTACTCTCAACTGCCGCGCAAACTTTAT | 1545 |
| DB | 1301 | AspIleGlnHisProAspThrTyrAspSerGlyMetGlnSerAsnThrValHisTyr | 1320 |
| QY | 1546 | AGTTTACTCAAGCATCTTAAACTCTGCCATAGCAGATTATCTTCACTATCTTATCAT | 1605 |
| DB | 1321 | AlaLeuMetThrSerLeuSerLeuAlaLeuAspAsn-----AsnTyrTyrIle--- | 1336 |
| QY | 1606 | CCAAAGGTGCTAGGATAGATGTTTCTTATCAATGATGTTATGATGCTCTTATCAGCA | 1665 |
| DB | 1337 | -----ThrGlnLeuAspIleSer-----SerAlaTyr-----LeuTyrAlaAsp | 1349 |

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

| Result No. | Score | Query | | DB | ID | Description |
|------------|--------|-------|--------|----|-------------|--------------------|
| | | Match | Length | | | |
| 1 | 3413 | 44.4 | 739 | 1 | JJAZ2 HUMAN | Q15022 homo sapien |
| 2 | 3332 | 43.3 | 741 | 1 | JJAZ2 MOUSE | Q80u70 mus musculu |
| 3 | 1127.5 | 14.7 | 900 | 1 | SUI12 DROME | Q9njg9 drosophila |
| 4 | 661 | 8.6 | 243 | 1 | JA21 HUMAN | Q86vz6 homo sapien |
| 5 | 661 | 8.6 | 243 | 1 | JA21 MOUSE | Q80zq5 mus musculu |
| 6 | 224 | 2.9 | 631 | 1 | VRN2 ARATH | Q816y4 arabidopsai |
| 7 | 217.5 | 2.8 | 440 | 1 | VRN2 ARATH | Q816y4 arabidopsai |
| 8 | 211 | 2.7 | 692 | 1 | F1S2 ARATH | Q8W5b1 arabidopsai |
| 9 | 176.5 | 2.3 | 1142 | 1 | G1N4 YEAST | Q9ntn9 arabidopsai |
| 10 | 164.5 | 2.1 | 1453 | 1 | Y373 BOVIN | Q12263 saccharomyc |
| 11 | 164.5 | 2.1 | 1875 | 1 | MLP1 YEAST | Q9tu23 bos taurus |
| 12 | 159.5 | 2.1 | 912 | 1 | UBP3 YEAST | Q02455 saccharomyc |
| 13 | 157 | 2.0 | 1308 | 1 | PCP1 SCHEPO | Q01477 saccharomyc |
| 14 | 155.5 | 2.0 | 1755 | 1 | YJ27 YEAST | Q92351 schizosacch |
| 15 | 153 | 2.0 | 707 | 1 | SFQ2 HUMAN | P47098 saccharomyc |
| 16 | 152 | 2.0 | 1755 | 1 | YJ29 YEAST | P23246 homo sapien |
| 17 | 152 | 2.0 | 1357 | 1 | SPOF SCHEPO | P47100 saccharomyc |
| 18 | 151 | 2.0 | 1328 | 1 | YMT5 YEAST | Q10411 schizosacch |
| | | | | | | Q04214 saccharomyc |

Db 312 AspGlyGluTyrGluValAlaMetGlnGluMetGluGluCysProIleSerIleValSerGly 331
 QY 1156 GCAACATGGAGACTATCTTTCATGGGAAGAGCTCCCTCATTCGAAACATTTCTCAG 1215
 Db 332 AlaThrTrpGluThrIleLeuAspGlyArgLeuProPheGluThrPheSerGln 351
 QY 1216 GGACCTACCTTTCAGTTCACTCTTCGTTGGACAGGAGACCAATGATAAATCTACGGCT 1275
 Db 352 GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspIleSerThrAla 371
 QY 1276 CTTATTCGCAACCTCTTCGCCACTAGAAATTCAGAGAGTCTCCATCAGAGAAACAGCCT 1335
 Db 372 ProIleAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro 391
 QY 1336 GGTTCAGTTAAACCTTAACTACTTGTCTGTTAAAGAAATCTTCACTACAGATCTACAA 1395
 Db 392 GlySerValLysProThrGlnThrIleAlaValLysGluSerLeuThrAspIleGln 411
 QY 1396 ACAAGAAAGAAAGATCTCCAAATGAAACCCGACAAATTAAGAAATTTATTCAG 1455
 Db 412 ThrArgLysGluLysAspThrProAsnGluAsnArgGlnLysLeuArgIlePheTyrGln 431
 QY 1456 TTTCTCTATAACAACAATACAGCAACAACCTGAGCAAGCAAGATGACTGCATTCGCCT 1515
 Db 432 PheLeuTyrAsnAsnThrArgGlnGlnThrGluAlaArgAspLeuHisCysPro 451
 QY 1516 TGGTGTACTCTGAATCGCGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCTGCCAT 1575
 Db 452 TrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisLeuLysCysHis 471
 QY 1576 ACCAGATTATCTTCAACTATCTTTATCATCAAAAGTGTAGATAGATGTTCTTATC 1635
 Db 472 SerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIle 491
 QY 1636 AATGAGTGTATGATGGCTCCATCAGGAATCTCAGGATATTCATCGCAACTCGGA 1695
 Db 492 AsnGluCysTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGly 511
 QY 1696 TTTGTTTATGTCGCAACGAGCATGTTAAGAACACCTTATCACACATATCTTGTGTC 1755
 Db 512 PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys 531
 QY 1756 AGGCCAAACGCAACAAAGCAGCATCTGNAATTTCTGATCTGAATCTGAAGATGGGAAGTA 1815
 Db 532 ArgProLysArgThrLysAlaSerMetSerGluPheLeuGluSerGluAspGlyGluVal 551
 QY 1816 GAACAGCAAAACATATAGTAGTGGCCACCAATCTCTCTATTTCCATAGTACACTGC 1875
 Db 552 GluGlnArgThrTyrSerSerGlyHisAsnArgLeuTyrPheHisSerAspThrCys 571
 QY 1876 TTACTCTCCGTCACAGAAATGGAAGTAGATAGATGAAAGATGATCTGATGG 1935
 Db 572 LeuProLeuArgProGlnGluMetGluValAspSerGluAspGluLysAspProGluTrp 591
 QY 1936 CTAAGAGAAAACCATACACAAATTCGAAGATTTCTGATGTTTAAAGAGGAGAGAAA 1995
 Db 592 LeuArgGluLysThrIleThrGlnIleGluPheSerAspValAsnGluGlyGluLys 611
 QY 1996 GAAGTGATGAACCTCGAAATCTCCATGTCATGAGCATGGGTTTATTCGTGACAAATCA 2055
 Db 612 GluValMetLysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 631
 QY 2056 ATGAATCATGCTGTATCTGTTCTAGAAATTTATGACAGAGAAATTAATTAAGAGAT 2115
 Db 632 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleLeuLysAsn 651
 QY 2116 TTATGTCGAAACTTCATCTTCATCTAGTCAGCATGCTGATCTTAAATCTTATAGCATA 2175
 Db 652 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 671
 QY 2176 ATGTCAATAGATAAGCTGTTACCAAGCTCGTGAATTCGACGCAAAATTAAGAAAGGG 2235

Db 672 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGluLysGly 691
 QY 2236 GAATCTCTCCCTCCCTCAAAACGAGAAATACTAGAGAAACAAAATCGGACAGCAATGGA 2295
 Db 692 GluSerAlaSerProAlaAsnGluGluIleThrGluGlnAsnGlyThrAlaAsnGly 711
 QY 2296 TTTAGTCAAAATTAACCTCAAAAGAGAAAGCTTTGGAAACACATAGTGTCTCAGGGTTTCA 2355
 Db 712 PheSerGluLeuAsnSerLysGluLysAlaLeuGluThrAspSerValSerGlyValSer 731
 QY 2356 AACAGAGCAAAAACAAAACATC 2379
 Db 732 LysGlnSerLysLysGlnLysLeu 739
 RESULT 2
 JUAZ_MOUSE
 ID JUAZ_MOUSE STANDARD; PRT; 741 AA.
 AC Q80U70; Q80V10; Q99L07;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Polycarb protein Suz12 (Suppressor of zeste 12 protein homolog).
 GN SUZ12 OR KIAA0150 OR D11ERTD530E.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Brain;
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
 RA Nakajima D., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
 RT cDNAs identified by screening of terminal sequences of cDNA clones
 RT randomly sampled from size-fractionated libraries.";
 RL DNA Res. 10:35-48(2003).
 RN [2]
 RC SEQUENCE OF 365-741 FROM N.A.
 RP STRAIN=FVB/N; TISSUE=Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Polycarb group (PcG) protein. PcG proteins act by
 CC forming multiprotein complexes, which are required to maintain the
 CC transcriptionally repressive state of homeotic genes throughout
 CC development. PcG proteins are not required to initiate repression,
 CC but to maintain it during later stages of development. They
 CC probably act via the methylation of histones, rendering chromatin
 CC heritably changed in its expressibility. Component of the PC2
 CC complex, which methylates lys-9 and lys-27 residues of histone H3
 CC (by similarity).
 CC -!- SUBUNIT: Component of the PC2 complex, composed of EBD, EZH2,
 CC SUZ12/JUAZ1, RBBP4 and RBBP7. This complex is distinct from the

CC
CC PRC1 complex, which contains many other PcG proteins like Bmi1,
CC CBX4, CBX8, PHC2, PHC3, SCML1, KIN01 and RNF2 (By similarity).
CC SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the VEPS (VRN2-EMF2-FIS2-SU(2)12) family.
CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to an
CC erroneous RNA joining and initiations.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

| | |
|----|---|
| CC | EMBL; AK122213; BAC65495.1; ALT_SEQ. |
| DR | EMBL; BC003922; AAH03922.1; - |
| DR | EMBL; BC051099; AAH51099.1; - |
| DR | MGD; MGI:1261758; D11Erttd5306. |
| DR | InterPro; IPR007087; Znf_C2H2. |
| DR | PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1. |
| DR | PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE NEG. |
| KW | Transcription regulation; Repressor; Chromatin regulator; |
| KW | Nuclear protein; Metal-binding; Zinc; Zinc-finger. |
| FT | ZN_FING 450 473 C2H2-TYPE. |
| FT | DOMAIN 565 641 VFEF-BOX. |
| FT | DOMAIN 7 50 GLY-RICH. |
| FT | DOMAIN 29 69 ALA-RICH. |
| SO | SEQUENCE 741 AA; 83025 MW; 4A6B6E5D24FAEF2 CRC64; |

| | | |
|--|----------|-----------------|
| Alignment Scores: | | |
| Pred. No.: | 6 3e-227 | 741 |
| Score: | 332.00 | 630 |
| Percent Similarity: | 98.161% | Conservative: 9 |
| Best Local Similarity: | 97.22% | Mismatches: 9 |
| Query Match: | 43.33% | Indels: 0 |
| DB: | 1 | Gaps: 0 |
| US-09-874-162A-7 (1-4409) x JJAZ MOUSE (1-741) | | |

| | | | |
|----|------|---|------|
| Qy | 916 | TTTAGAGTACTCGTCCAGGAGACAGAGATTAAATGGAAATGATTAATCGAGAAACCAAT | 975 |
| Db | 254 | PheArgValThrArgProGlyArgArgGluPheAsnGlyMetIleAsnGlyGluThrAsn | 273 |
| Qy | 976 | GAATAATTATGATCTCAATGAAGAGCTTCCAGCCAGAGAAACGAAATCCTGAGGATGGG | 1035 |
| Db | 274 | GluAsnIleaspValSerGluGluLeuProAlaIaArgArgAsnArgGluaspGly | 293 |
| Qy | 1036 | GAAGAAGACATTTGTCGCAAAATGACAGTATTTGTATATAAAAACAGGCGCTTACAGCTTTTA | 1095 |
| Db | 294 | GluLysThrPheValAlaGlnMetThrValPheAspLysAsnArgArgLeuGlnLeuLeu | 313 |
| Qy | 1096 | GATGGGGAATATCAAGTTCAGCTAGCCATCGAGGAAATGGAGAATGTCCAATAAGCAAGAAAAAG | 1155 |
| Db | 314 | AspGlyGluTyrGluValAlaMetGlnGluMetGluGluCysProIleSerIysLysArg | 333 |
| Qy | 1156 | GCACATGGAGAGACTATTTGTATGGGAAGAGCGTCCTCCATTGCAAAACATTTTCTCAG | 1215 |
| Db | 334 | AlaThrTrpGluThrIleLeuaspGlyLysArgLeuProPheGluThrPheSerGln | 353 |
| Qy | 1216 | GCACCTACGTTGCAGTTCACCTCTTCGTCGACAGGAGACCAATGATAAATCTACCGCT | 1275 |
| Db | 354 | GlyProThrLeuGlnPheThrLeuArgTrpThrGlyGluThrAsnAspLysSerThrAla | 373 |
| Qy | 1276 | CCATATTGCCAAACCTCTTGCCACTAGAAATTCAGAGAGTCTCCATCAGAGAAACAAGCCT | 1335 |
| Db | 374 | ProValAlaLysProLeuAlaThrArgAsnSerGluSerLeuHisGlnGluAsnLysPro | 393 |
| Qy | 1336 | GGTTCAGTTAAACCTACTCAACTATTCGTGTAAAGAATCAATGACTACAGATCTACAA | 1395 |
| Db | 394 | GlySerValLysProAlaGlnThrIleAlaValLysGluThrLeuThrThrGluLeuGln | 413 |
| Qy | 1396 | ACAAGAAAAGAAAGATACTCCAAATGAAAAACCGACAAAAATTAAGAATATTTTATCAG | 1455 |
| Db | 414 | ThrArgLysGluLysAspAsnSerAsnGluSerArgGlnLysLeuArgIlePheTyrGln | 433 |
| Qy | 1456 | TTTCTCTATAACAACATACAAGGCCAACAACTGAAGCAAGAGATGACCTGCAATGCCCT | 1515 |
| Db | 434 | PheLeuTyrAsnAsnAsnThrArgGlnGlnThrGluAlaArgAspAspLeuHisCysPro | 453 |
| Qy | 1516 | TGCTGTACTCTGAACGCGCAACCTTATAGTTTACTCAAGCATCTTAACTCTGCCAT | 1575 |
| Db | 454 | TrpCysThrLeuAsnCysArgLysLeuTyrSerLeuLeuLysHisIleuLysLeuCysHis | 473 |
| Qy | 1576 | AGCAGATTTTATCTTCAACTATGTTTATCATCCAAAAGGTGCTAGGATAGATGTTTCTATC | 1635 |
| Db | 474 | SerArgPheIlePheAsnTyrValTyrHisProLysGlyAlaArgIleAspValSerIle | 493 |
| Qy | 1636 | AATGAGTGTATGATGGCTCCTATGACAGAAATCCTCAGGATATTCATGCCACCTGGA | 1695 |
| Db | 494 | AsnGluCysTyrAspGlySerTyrAlaGlyAsnProGlnAspIleHisArgGlnProGly | 513 |
| Qy | 1696 | TTTGCTTTTATGTCGCAACGACAGTAAAGAAACACCTTATCACATATCTCTGTGTGC | 1755 |
| Db | 514 | PheAlaPheSerArgAsnGlyProValLysArgThrProIleThrHisIleLeuValCys | 533 |
| Qy | 1756 | AGGCCAAAACGAACAAAAGACATGTCTGAATTTCTGAATCTCAAGATGGGGAAGTA | 1815 |
| Db | 534 | ArgProLysArgThrIysAlaSerMetSerGluPheLeuGluSerGluaspGlyGluVal | 553 |
| Qy | 1816 | GRACACGAAGACATATAGTAGTGCCCAATTCGCTGTTTTCATAGTATGATCCTGC | 1875 |
| Db | 554 | GlnGlnGlnArgThrTyrSerSerGlyHisAsnArgLeuTyrPheHisSerIaspThrCys | 573 |
| Qy | 1876 | TTTACCTCTCCGTCACAGAATAATGGAAGTAGATAGTGAAGATGAAAAGGATCCTGAATCG | 1935 |
| Db | 574 | LeuProLeuArgProGlnGluMetGluValAspSerGluaspGluLysAspProGluTrp | 593 |
| Qy | 1936 | CTAAGAGAAAAACCAATTACAAAAATGGAAGTTCCTGATGCTTAATCAAGAGAGAAA | 1995 |
| Db | 594 | LeuArgGluLysThrIleThrGlnIleGluGluPheSerAspValAsnGluGlyGluLys | 613 |
| Qy | 1996 | GAAGTGAATGAAACTCTGGAATCTCCATGTGCATGAAGCATGGGTATTTCGTCAACATCAA | 2055 |

```

Db      614 GluValMetCysLeuTrpAsnLeuHisValMetLysHisGlyPheIleAlaAspAsnGln 633
QY      2056 ATGAATCATGCTGTATGCTGTTGTAGAAAATTATGGACAGAAATAATTAAGAAGAAAT 2115
Db      634 MetAsnHisAlaCysMetLeuPheValGluAsnTyrGlyGlnLysIleLeuLysLysAsn 653
QY      2116 TTATGTCGAAATTCATGCTTCATCTAGTCAGCATGCGATCTTAAATCTTATTAGATA 2175
Db      654 LeuCysArgAsnPheMetLeuHisLeuValSerMetHisAspPheAsnLeuIleSerIle 673
QY      2176 ATGTCAATAGATAAAGCTGTTACCAAGCTCCGTGAATGCGACAAAATATGAAAAGGGG 2235
Db      674 MetSerIleAspLysAlaValThrLysLeuArgGluMetGlnGlnLysLeuGlnLysGly 693
QY      2236 GAATCTGCTTCCCTCGAACGAGAAATACTAGGAACAAATAATGGACAGCAAAATGGA 2295
Db      694 GluSerAlaThrProSerAsnGluGluIleAlaGluGluGlnAsnGlyThrAlaAsnGly 713
QY      2296 TTTAGTGAATTAACCTCAAAAGAGAAAGCTTTGGAACAGATAGTGTCTCAGGGGTTTCA 2355
Db      714 PheSerGluThrAsnSerLysGlnLysAlaLeuGluThrAspGlyValSerGlyValPro 733
QY      2356 AAACAGACAGAAAACAAACAAATC 2379
Db      734 LysGlnSerLysLysGlnLysLeu 741
RESULT 3
SUI2 DROME
ID Q9UG9; Q8T9D8; Q9VW55; PRT; 900 AA.
DC 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Polcomb protein Su(z)12 (Suppressor 12 of zeste protein).
GN Su(z)12 OR CG8013.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), CHARACTERIZATION, AND MUTANT
RX SU(z)12-2.
RX MEDLINE=21430867; PubMed=11546753;
RA Birve A., Sengupta A.K., Beuchle D., Larsson J., Kennison J.A.,
RA Rasmuson-Lestander A., Mueller J.
RT "Su(z)12, a novel Drosophila Polcomb group gene that is conserved in
RT vertebrates and plants.";
RL Development 128:3371-3379 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazee R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

```

```

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Rue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Stadler A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RX MEDLINE=2445609; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537589;
RA Scapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
RN [5]
RP IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAFI AND ESC,
RP AND METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
RX MEDLINE=22296673; PubMed=12408863;
RA Cermin B., Melfi R., McCabe D., Seitz V., Imhof A., Pirrotta V.;
RT "Drosophila enhancer of Zeste/ESC complex has a histone H3
RT methyltransferase activity that marks chromosomal Polycomb sites.";
RL Cell 111:185-196 (2002).
RN [6]
RP IDENTIFICATION IN A ESC/E(Z) COMPLEX WITH E(Z); CAFI AND ESC, AND
RP METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
RX MEDLINE=22296674; PubMed=12408864;
RA Mueller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A.,
RA Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A.;
RT "Histone methyltransferase activity of a Drosophila Polycomb group
RT repressor complex.";
RL Cell 111:197-208 (2002).
RN [7]
RP -1- FUNCTION: Polycomb group (PcG) protein. PcG proteins act by
RP forming multiprotein complexes, which are required to maintain the
RP transcriptionally repressive state of homeotic genes throughout
RP development. PcG proteins are not required to initiate repression,
RP but to maintain it during later stages of development. They
RP probably act via the methylation of histones, rendering chromatin
RP heritably changed in its expressibility. Component of the Esc/E(z)
RP complex, which methylates Lys-9 and Lys-27 residues of histone H3.
RP Despite the presence of a zinc-finger, it does not bind directly
RP to DNA, the Esc/E(z) complex being probably recruited to DNA by
RP Pto. The Esc/E(z) complex is necessary but not sufficient to
RP recruit a functional PcG repressive complex that represses target
RP genes, suggesting that the recruitment of the distinct PRC1
RP complex is also required to allow a subsequent repression.
RN [8]
RP -1- SUBUNIT: Component of the Esc/E(z) complex, composed of Esc, E(z),

```


Db 362 AsnAsnAsnAsnCysSerGlyLeuLysAsnGlySerGlyGlyAsnSerThrVal 381
 QY 1363 GCTGTTAAAGATCATCTAGATCTACAGATCTACAAACAGAAAAGAAAGGATACTCCAAT 1422
 Db 382 Cys-----LysThrThrPro--- 386
 QY 1423 GAAACCGACAAAATTAAGAAATATTTTATCAGTTCTCTATAACAACAATAACAGGCAA 1482
 Db 387 -----GluHisIleGlnIleValTyAsnPhemetyrSerAsnAsnThrArgGln 403
 QY 1483 CAAACTGAAGCAGAGATGACCTGCATTCCTCCCTTGGTGACTCTGAACTCGCGCAACTT 1542
 Db 404 GlnThrGluTyThrGlnGluLeuAsnCysProTyrCysGlyLeuAspCysLeuArgLeu 423
 QY 1543 TATAGTTTACTCAAGCATCTTAACTCTGCCATACAGATTTTATCTTCAACTATGTTAT 1602
 Db 424 TyrAlaLeuLeuLysHisLeuLysLeuCysHisAlaArgPheAsnPhethrTyArgPro 443
 QY 1603 CATCCAAAAGGTCTAGTAGATGTTTCTATCAATGAGTGTTATGATGGTCTCCATGCA 1662
 Db 444 AlaGlySerGlyAlaArgIleAspValThrIleAsnAspAlaTyAspGlySerTyrAla 463
 QY 1663 GGAATCTCCAGATATTCATCGCAACCTGGATTTGCTTTAGTCGCAAC---GGACCA 1719
 Db 464 GlySerProTyrAspLeuAlaGlyProSerGlySerSerPheAlaArgThrCysGlyPro 483
 QY 1720 GTTAAGAGAACACCTTATCACATATCTGTGTGCAGCCCAACCAACGACAAAGCAAGC 1779
 Db 484 ValArgArgThrSerValThrSerLeuMetValCysArgProArgArgGlnLysThrCys 503
 QY 1780 ATGCTCTGAATTTCTGAATCTGAAGATGGGGAAGTAGAACGACCAAGAACATATAGTAGT 1839
 Db 504 LeuAspGluPheLeuGluLeuAspGluAspGluIleSerAsnGlnArgSerTyrIleThr 523
 QY 1840 GCCCAATCGTCTGTTTTCATAGTAGTACCTGCTTACCTCCCGTCCACAGAATG 1899
 Db 524 GlyHisAsnAsnGleuTyHisHisThrGluThrCysLeuProValHisProLysGluLeu 543
 QY 1900 GAAGTAGTAGTAGAGATCAAGATCTCTGAATGGCTAAGAGAAAACCATACACAA 1959
 Db 544 AspileAspSerGluGlyGluSerAspProLeuTrpLeuArgGlnLysThrIleGlnMet 563
 QY 1960 ATTGAAGATTTCTGATGTTAATGAAGAGAGAGAAAGTAGTGAACACTCGGAATCTC 2019
 Db 564 IleAspGluPheSerAspValAsnGluGlyGlyLeuMetLysLeuTrpAsnLeu 583
 QY 2020 CATGTCATGAAGCATGGGTTTATTGCTGACAAATCAATGAATCATGCTGATCTCTTT 2079
 Db 584 HisValMetArgHisGlyPheValGlyAspCysGlnLeuProIleAlaCysGluMetPhe 603
 QY 2080 GTAGAAATATGACAGACAAAATTAATGAAGAAATTTATGTCGAAACTTCATGCTTCAT 2139
 Db 604 LeuAspAlaLysGlyThrGluIleValArgLysAsnLeuTyArgAsnPhelleLeuHis 623
 QY 2140 CTAGTCATGATGATGCTTAAATCTTATTAGCATATGTCATAGATAAGCTGTGTTACC 2199
 Db 624 MetCysSerLeuPheAspTyGlyLeuIleAlaAlaGluThrValTyLysThrValGln 643
 QY 2200 AAGTCCTCGTAATGACGACAAAATTAAGAAAGGGGGAATCT----- 2241
 Db 644 LysLeuGlnGlyLeuLeuSerLysTyAlaAlaGlyGlnGluMetGlnArgGlnArg 663
 QY 2241 ----- 2241
 Db 664 GluGluGlnLeuLysTyTrpLeuAspValGlyMetHisLysLysGlnGluAspProLys 683
 QY 2242 -----GCTTCCCTCCAAACGAAAGAAATAAATCACTGAA----- 2271
 Db 684 ThrLeuLysSerProGlnLysProAlaProProAlaAspGlnAlaSerThrSerSerAla 703
 QY 2271 ----- 2271
 Db 704 SerThrSerGlySerGlySerGlySerSerMetGlnProProLysArgMetProAla 723

QY 2272 -----GAACAAATGGACAGCAAT 2292
 Db 724 HisLeuLysArgGlySerAlaAlaSerSerProGlyValGlnSerLysGlyThrGluAsn 743
 QY 2293 GCATTAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2352
 Db 744 GlyThrAsnGlySerAsnSer-----SerSerSerAsnSerLysAsnVal 758
 QY 2353 TCAAAACAGACAGCAAAACCAAAACCAAACTC 2379
 Db 759 AlalysLysSerAlaAspGlnProLeu 767
 RESULT 4
 JAZ1 HUMAN
 ID JAZ1_HUMAN
 AC Q86VZ6; Q8N3L7; STANDARD; PRT; 243 AA.
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Juxtaposed with another zinc finger protein 1.
 GN JAZF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain, and Testis;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallory S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalón D.K., Munz D.C., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Amalgam;
 RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP DISEASE, AND CHROMOSOMAL TRANSLOCATION WITH SUZ12.
 RX MEDLINE=21265470; PubMed=11371647;
 RA Koontz J.I., Soreng A.L., Nucci M., Kuo F.C., Pauwels P.,
 van Den Berghe H., Cln P.D., Fletcher J.A., Sklar J.;
 RT "Frequent fusion of the JAZF1 and JAZ1 genes in endometrial stromal
 tumors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:6348-6353(2001).
 CC -!- FUNCTION: Potential transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear protein (Potential).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC Name=2;
 CC IsoId=Q86VZ6-1; Sequence=Displayed;
 CC IsoId=Q86VZ6-2; Sequence=VSP_007755;
 CC Note=Sequence incomplete. NO experimental confirmation
 available;

CC -!- DISEASE: Endometrial stromal tumors can be caused by a chromosomal
 CC rearrangement involving JAZF1 and SUZ12. A t(7;17)(p15;q21)
 CC chromosomal translocation involving JAZF1 and SUZ12 generates the
 CC JAZF1-SUZ12 oncogene consisting of the N-terminus part of JAZF1
 CC and the C-terminus part of SUZ12. The translocation is frequently
 CC found in all cases of endometrial stromal tumors, except in
 CC endometrial stromal sarcomas, where it is rarer.
 CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; BC042441; AAH42441.1; -;
 CC EMBL; AL834234; CAD38912.1; -;
 CC MIN; 606246; -;
 CC InterPro: IPR007087; Znf_C2H2;
 CC Pfam; PF00096; Znf_C2H2; 3;
 CC SMART; SM00335; Znf_C2H2; 3;
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2; FALSE NEG.
 CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; ZINC_FINGER_C2H2; 3;
 CC Transcription regulation; Nuclear protein; Metal-binding; Zinc;
 CC Zinc-finger; Chromosomal translocation; Proto-oncogene;
 CC Alternative splicing.
 CC FT ZN_FING 12 37 C2H2-TYPE 1.
 CC FT ZN_FING 173 198 C2H2-TYPE 2.
 CC FT ZN_FING 208 230 C2H2-TYPE 3. (DEGENERATE)
 CC FT SITE 129 130 BREAKPOINT FOR TRANSLLOCATION TO FORM
 CC FT VARSPLIC 1 38 JAZF1-SUZ12 ONCOGENE.
 CC FT MTGFAASFFSNTCTFCGCGLHPTFLADLIEHEDNHI ->
 CC FT YFVLGSLIYDAKSKNRCNVFIT (in isoform 2).
 CC FT /FTId=VSP 007755.
 CC FT D -> G (IN REF. 1).
 CC FT CONFLICT 133 133 27079 MW; 878A6EE3D82C0588 CRC64;
 CC SEQUENCE 243 AA; 27079 MW; 878A6EE3D82C0588 CRC64;

Alignment Scores:
 Pred. No.: 6,3e-39 Length: 243
 Score: 661.00 Matches: 128
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.60% Indels: 0
 DB: 1 Gaps: 0

US-09-874-162a-7 (1-4409) x JAZ1_HUMAN (1-243)
 QY 52 ATGACAGCATCGCGCGCTCTCTCTCCATACCTCGGATCGGGGCTCGGGA 111
 DB 1 MetThrGlyLeuAlaAlaSerPheSerThrCysArgPheGlyGlyCysGly 20
 QY 112 CTCACACTCCCGCCGACCTCATCGAGCATCGAGCAACACCATCATGATA 171
 DB 21 LeuHisPheProThrLeuAlaAspLeuGluHisLeuGluAspAsnHisLeuAspThr 40
 QY 172 GATCCACGGTGTAGAAAACAGAAATACAGCAGCAGCACTATGTGCGCTGAGTTAC 231
 DB 41 AspProArgValLeuGluGlyGluLeuGlnGlnProThrThrValAlaLeuSerTyr 60
 QY 232 ATAAATAGATTTCATCAGATGCTCCCGCCGAGAGGAGTCCCTAAAGAAGATT 291
 DB 61 IleAsnArgPheMetThrAspAlaAlaArgGluGlnGluSerLeuLysLysLysile 80
 QY 292 CAGCCGAGAGCTCTGCTGACTCTGTGCGAGTCTGAGGGAATGTTCACATCCC 351
 DB 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100
 QY 352 CCACGCCACAGCAGGTGGAGGCTTACTCCCGCTGACCCACCATCACCCCTCTCT 411
 DB 101 ProArgHisSerSerGlySerLeuThrProProValThrProProLeuThrProSerSer 120

QY 412 TCATTCCGACGACGACATCCGACA 435
 DB 121 SerPheArgSerSerThrProThr 128
 RESULT 5
 JAZ1_MOUSE
 ID JAZ1_MOUSE STANDARD; PRT; 243 AA.
 AC Q80ZQ5; Q8BLQ8; Q8BNW4;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Juxta-posed with another zinc finger protein 1.
 GN JAZF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A. (ISOFORM 1).
 RP STRAIN=CS7BL/6J; TISSUE=Brain cortex, and Pancreas;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,
 RA Baladrelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Forrest A., Frazer K.S.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konggava A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
 RA Ravasi T., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan A., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 [2]
 SEQUENCE FROM N.A. (ISOFORM 2).
 RP TISSUE=Testicle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max A.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Raheij J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Butterfield Y.S., Krzywinski M.I., Skalska U., Smalios D.E.,
 RA Bunker A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

-!- FUNCTION: Potential transcription factor.

-!- SUBCELLULAR LOCATION: Nuclear protein (Potential).

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=1;

isoId=Q80ZQ5-1; Sequences=Displayed;

Name=2;

isoId=Q80ZQ5-2; Sequences=VSP 007756;

Note=No experimental confirmation available;

-!- SIMILARITY: Contains 3 C2H2-type zinc fingers.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; AK043767; BAC31649.1; -.

DR EMBL; AK050474; BAC34275.1; -.

DR EMBL; BC048577; AAH48577.1; -.

DR InterPro; IPR007087; Znf C2H2.

DR Pfam; PF00096; zf-C2H2; 3.

DR SMART; SM00355; Znf C2H2; 3.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; FALSE NEG.

KW Transcription regulation; Nuclear protein; Metal-binding; Zinc;

XW Zinc-finger; Alternative splicing.

FT ZN_FING 12 37 C2H2-TYPE 1.

FT ZN_FING 173 198 C2H2-TYPE 2.

FT ZN_FING 208 230 C2H2-TYPE 3 (DEGENERATE).

FT VARSLIC 64 64 Missing (in isoform 2).

FT /FTID=VSP_007756.

FT CONFLICT 211 211 R -> L (IN REF. 1; BAC34275).

SQ SEQUENCE 243 AA; 27097 MW; 878A6EB3D82C0348 CRC64;

Alignment Scores:

| Pred. No.: | Score: | Length: | Matches: |
|------------------------|---------|---------|----------|
| 6.3e-39 | 243 | | |
| 661.00 | 128 | | |
| Percent Similarity: | 100.00% | | |
| Best Local Similarity: | 100.00% | | |
| Query Match: | 8.60% | | |
| DB: | 1 | | |

US-09-874-162A-7 (1-4409) x JAZ1_MOUSE (1-243)

QY 52 ATGACAGGATCCCGCCGCTCTTCTTCTCCATACCTGCGGATTCGGGGCTGCGGA 111

Db 1 MetThrGlyIleAlaAlaSerPhePheSerAsnThrCysArgPheGlyGlyCysGly 20

QY 112 CTCGACTTCCACCTGCGCCACCTCATCGAGCATCGAGGACACACATCGATACA 171

Db 21 LeuHisPheProThrLeuAlaAspLeuIleGluHisIleGluAspAsnHisIleAspThr 40

QY 172 GATCCACGGGTTTATAGAAAAACAAGATTACAGCAGCAACCTATGTTGCCCTGAGTTAC 231

Db 41 AspProArgValLeuGlyGlnGluLeuGlnGlnProThrTyValAlaLeuSerTyr 60

QY 232 ATAATAGATTATGACAGATCTCTCCCGCAGAGCGAGAGTCCCTAAAGAGAGATT 291

Db 61 IleAsnArgPheMetThrAspAlaAlaArgGluGlnGluSerLeuLysLysLysIle 80

QY 292 CAGCCGAGCTCTCGCTGACTCTGCTCCAGCTCAGTCTCGAGGGAATGTCCTCACTCC 351

Db 81 GlnProLysLeuSerLeuThrLeuSerSerValSerArgGlyAsnValSerThrPro 100

QY 352 CCACGCCACAGAGTGAAGCTTACTCCCGCGTGACCCACCCATCACCCCTCTCT 411

Db 101 ProArgHisSerSerGlySerLeuThrProProValThrProIleThrProSerSer 120

QY 412 TCATTCCGACGAGCACTCCGACA 435

Db 121 SerPheArgSerThrProThr 128

RESULT 6

EMF2_ARATH STANDARD; PRT; 631 AA.

ID EMF2_ARATH Q8L6Y4; Q93V59; Q9LU50; Q9LU51;

AC Q8L6Y4; Q93V59; Q9LU50; Q9LU51;

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Polycomb protein EMBRYONIC FLOWER 2.

GN EMF2 OR AF5G51230/AF5G51240 OR MWD22.18/MWD22.19.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC STRAIN=cv. Columbia;

RX MEDLINE=2158794; PubMed=11701882;

RA Yoshida N., Yanai Y., Chen L., Kato Y., Hiratsuka J., Miwa T.,

RA Sung Z.R., Takahashi S.,

RT "EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates

RT shoot development and flowering in Arabidopsis.";

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=20181125; PubMed=10718197;

RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,

RA Tabata S.,

RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence

RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC

RT clones.";

RL DNA Res. 7:31-63 (2000).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RC STRAIN=cv. Columbia;

RX MEDLINE=22954850; PubMed=14593172;

RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J., Cheuk R.P.,

RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.P.,

RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,

RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,

RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,

RA Arakawa T., Barh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,

RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,

RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,

RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,

RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,

RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,

RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,

RT "Empirical analysis of transcriptional activity in the Arabidopsis

RT genome.";

RL Science 302:842-846 (2003).

-!- FUNCTION: Polycomb group (PcG) protein. Involved in flowering processes by repressing unknown target genes and preventing reproductive development. PcG proteins act by forming multiprotein complexes, which are required to maintain the transcriptionally repressive state of homeotic genes throughout development. PcG proteins are not required to initiate repression, but to maintain it during later stages of development. They probably act via the methylation of histones, rendering chromatin heritably changed in its expressibility.

-!- SUBUNIT: In plants, PcG complexes are probably composed of a member of the EZ family (CLF or MEA), FIE, and a member of the VFP family (FIS2, VRN2 or EMF2) (By similarity).

-!- SUBCELLULAR LOCATION: Nuclear (Probable).

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=1;

isoId=Q8L6Y4-1; Sequence=Displayed;

Name=2;

Db 379 AspAspValAspProGlnGlnThrPhePheSerSerLysLysPheArgArg 398
QY 1705 -----AGTCGCAACGGACCA----- 1719
Db 399 ArgGlnLysSerGlnValArgSerArgGlnGlnProHisLysLysLeuGlyCysGlu 418
QY 1720 -----GTTAAGAGAACCTATACATATCTGTGTGTCAGGCCAAACGA----- 1767
Db 419 ValLeuAspLysThrAspAlaHisSerValArgSerGluLysSerArgIleProPro 438
QY 1767 ----- 1767
Db 439 GlyLysHisTyrGluArgIleGlyAlaGluSerGlyGlnArgValProProGlyThr 458
QY 1768 -----ACAAAGCA 1776
Db 459 SerProAlaAspValGlnSerCysGlyAspProAspTyrValGlnSerIleAlaGlySer 478
QY 1777 ACATGCTGCAATTCCTGAATCTGAAGTGGGAGTAGACACAA-----AGACAA 1830
Db 479 ThrMetLeuGlnPheAlaLysThrArgLysIleSerIleGluArgSerAspLeuArgAsn 498
QY 1831 TATAGTAGTGGCCACAATCGTCTGTATTTCCATAGTAGTACCTGCTTACCTCTCGTCCA 1890
Db 499 ArgSerLeuLeuGlnLysArgGlnPhePheHisSerHisArgAlaGlnProMetAlaLeu 518
QY 1891 CAAGAAATG-----GAAGTAGATGAGATGAAGATGAAGATGCTGATGCTGAAGAA 1944
Db 519 GluGlnValLeuSerAspArgAspSerGluAspGluValAlaAspValAlaAspPhe 538
QY 1945 AAAACCATTACACAAATTCAGAGTTTCTGTATGTTAATGAGGAGAGAGAGAGATGATG 2004
Db 539 GluAspArgArgMetLeuAspPheValAspValThrLysAspGluLysGlnMetMet 558
QY 2005 AAATCTGGAATCTCATGATGATGAGCATGAGTGGTTTATGCTGCAATCAATCAATCAT 2064
Db 559 HisMetTrpAsnSerPheValArgLysGlnArgValLeuAlaAspGlyHisIleProTrp 578
QY 2065 GCCTGTATGCTCTTCTGACAAATATGACAGAAATAATTAAG---AAGNATTTATGT 2121
Db 579 AlaCysGluAlaPheSerArgLeuHisGlyProIleMetValArgThrProHisLeuIle 598
QY 2122 CGAAATCTCATGCTCATCTAGTCAGCATGATGCTTAACTTATTAGCATATGTC 2181
Db 599 TrpCysTrpArgValPheMetValLysLeuTrpAsnHisGlyLeuLeuAspAlaArgThr 618
QY 2182 ATAGATAAGCTGTACCAAGCTCCGTGAAATGCAG 2217
Db 619 MetAsnAsnCysAsnThrPheLeuGluGlnLeuGln 630
RESULT 7
ID VRN2_ARATH STANDARD; PRT; 440 AA.
AC Q8W5B1; O23524; Q8W5B2; Q94CF5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Polyclonal group protein VERNALIZATION 2.
GN VRN2 OR AT4G16845 OR FCALL23 OR DL4450W.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]_TaxID=3702;
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), FUNCTION, AND CHARACTERIZATION.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=21575875; PubMed=11719192;
RA Gendall A.R., Levy Y.Y., Wilson A., Dean C.;
RT "The VERNALIZATION2 (VRN2) gene mediates the epigenetic regulation of
RT vernalization in Arabidopsis.";
RL Cell 107:525-535 (2001).

RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=98121113; PubMed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H.M., Dean C.,
RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P.,
RA Wedler H., Wedler E., Wambutt R., Weitzneger T., Pohl T., Terryn N.,
RA Gielen J., Villarroel R., De Clercq R., van Montagu M., Lecharny A.,
RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S.,
RA Kouter P., Entian K.-D., Rieger M., Schaefer M., Funk B.,
RA Mueller-Auer S., Silvey M., James R., Monfort A., Pons A.,
RA Puigdomenech P., Douka A., Vouklatou E., Milioni D., Hatzopoulos P.,
RA Piravandi E., Obermaier B., Hilbert H., Duesterhoeft A., Moores T.,
RA Jones J.D.G., Eneva T., Palme K., Benes V., Rechmann S., Ansoorge W.,
RA Cooke R., Berger C., Delseny M., Voet M., Volckaert G., Mewes H.-W.,
RA Klosterman S., Schueller C., Chalwatzis N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana.";
RL Nature 391:485-488 (1998).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083489; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Breken M., Weljens J., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzneger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moolijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLeay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Ganderath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Feilber R.,
RA Schnabl S., Hillier R., Berger C., Monfort A., Casacuberta E.,
RA Chefidor F., Cooke R., Shear C., Cordes M., Abu-Threideh J.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse C.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Laetelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA Nelson J., Spiech J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshua C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shchay N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M.A., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777 (1999).
[4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=cv. Columbia;

Db 469 ArgAlaSerValIleAsnValGluLysAsnGluLysArgAlaSerIlePheSerThrThr 488
 QY 586 GAGAAATGAAGAGAGAGCAAGATCT----- 612
 Db 489 LysLysAsnLysArgSerArgSerIleLysArgMetSerLeuIleProSerMetLys 508
 QY 613 ---CATAGCTGTGAGCTATTCCTGAGCTACGTTTACTGTTTCTTCCACAAAATGAT 669
 Db 509 ArgGluSerValThrThrLysLeuMetSerThrThrLysLeuAlaGluAspAsp 528
 QY 670 AAGCCATCACCACAACTCAGAAATGAACAAATCTCTTACCCCTGGAGTCTGCTGTG 729
 Db 529 TrpGluTyrlleGluLysGluThrLysArgThrSerSerAsnPheAlaThrLeuIleAsp 548
 QY 730 AAAGTT-----TGCCACAAAACAAAGAGATGAAGTTGT 765
 Db 549 GluIlePheGluTyrlleGluLysGluThrLysGluLysGluLysGluLysGluLys 568
 QY 766 CCAATAAGGCAAGTCCACAGGT----- 789
 Db 569 LysValArgGluAlaLysAlaArgGluLysGluLysGluLysGluLysGluLys 588
 QY 789 ----- 789
 Db 589 LysGluArgAlaArgLysLeuLysGluLysGluLysGluLysGluLysGluLys 608
 QY 790 AAAAAGCAGGTCCTTG---ATTCTGACCTCAATCAA----- 825
 Db 609 LysLysGluIleGluIleAspIleSerAspLeuGluGluGluLysLysLysGlu 628
 QY 826 ---ACAAAACCCGAAATTTCCCTCCCTGCGATTTCCAGTAAGTAATTTGAACCTAGT 882
 Db 629 GluLysLeuAspGlyAsnIleArgSerIleSerAlaProMetGluAsnGluLysAsn 648
 QY 883 AACGCCATATGGTGAAGCTTACTCGTTGCTATTAGA-----GTG 924
 Db 649 IleAsnHisLeuGluValAlaPheAspAsnIleLeuArgArgArgAsnPheSerLeuGln 668
 QY 925 ACTGCTCCAGGAAGAAGAGAGTTTAATGGAATGATTAATGAGAAACCAATGAATATT 984
 Db 669 ThrArgProValSerArgLeuAspProGlyIleMetPheSerSerProThrGluGluVal 688
 QY 985 GATGTCAATGAAGAGCTCCAGCAGAGAAACAAATCGTAGGATGGGAAAAGACA 1044
 Db 689 SerProValGlu-----ProLysArgThrGluAsnGluArgLeuThrThrGluLys 706
 QY 1045 TTGTGTGCACAAATG-----ACAGTATTTGATAAAAACAGCGCG 1083
 Db 707 IleLeuGluThrIleArgSerLysPheLeuGlySerPheAsnIleAspLysGlu 726
 QY 1084 TTACAGCTTTTAGTGGGATATGAA-----GTAGCCATGCGAGGAATG---GAGAA 1134
 Db 727 LeuLysLeuSerLysMetGluTyrlleProSerIleAlaProGlnArgLeuSerGlu 746
 QY 1135 TGTCACAAATGAACAAAGAGACATGGAGACTATTCT-----GATGGGAAG 1185
 Db 747 ArgValValSerAspSerAspGlyTyrlleGluSerLeuIleLeuProLysAspGlyAsn 766
 QY 1186 AGGCTGCTCCATTCGAA-----ACATTTCTCAGGACCTACGTTGCGAGTTCACTCTT 1239
 Db 767 GlyValSerGlnLeuLysAspSerThrAlaThrAlaProValSerAspGlyArgLeu 786
 QY 1240 GCTTGGCAGGAGACCAATGATAAATCTACGGCTCTATGTCACAACTCTTGCCACT 1299
 Db 787 ArgLysIleSerGlu-----IleArgValProGlnPheThr 798
 QY 1300 AGAAATTCAGAGAGTCCATCAGGAAACAAAGCTGTTGCTGAGTAAACCTTCAAACT 1359
 Db 799 ArgLysSerArgHisPheSerGluSerAsnLysArgLeuSerVal-----LeuSerMet 816
 QY 1360 ATTGCTGTTAAGATCATTTACTACATCTACAAACAAAGAAAGAGATCTCCA 1419
 Db 817 TyrSerThrLysGluSerPheThrAsnLeuValAspIleLeuLysAsnGlyAsnLeuAsp 836

QY 1420 AATGAAACCGACAAAATTAAGAATATTTTATCAGTTTCTCTATAACAACATACAAG 1479
 Db 837 ValAsnAsnGlnGlnSerGlnArgIle----- 845
 QY 1480 CAACAAACTGAAGCAAGAGATGACCTGCATTCGCTTGGTGTACTCTCAACTCGGCAAA 1539
 Db 846 ProThrProArgSerAlaAspAsp----- 853
 QY 1540 CTTTATAGTTTACTCAACCATCTTAACTCGCCATAGCAGATTTATCTTCAACTATGTT 1599
 Db 854 -----SerGluPheLeuPheGlu----- 859
 QY 1600 TATCATCCAAAAGGTCTAGGATAGATGTTTCTATCAATAGTGTATGATGGCTCTAT 1659
 Db 860 -----ThrValAsnGlu-----GluAlaGluTyr 867
 QY 1660 GCAGAAATCCTCAGGATATTCATCGC----- 1686
 Db 868 ThrGlyAsnSerSerAsnAspGluArgLeuTyrAspValGlyAspSerThrIleLysAsp 887
 QY 1687 -----CAACCTGGATTGCT-----TTTACTCGCAACGCGACCACTTAAGAGA 1728
 Db 888 LysSerAlaLeuLysLeuAsnPheAlaAspArgPheAsnGlySerAsnGluAlaLysGln 907
 QY 1729 ACACCTATCACATATTTCTGTGTGTCAGGCGCAAAACGAAACAAAGCAGCATGCTGAA 1788
 Db 908 ThrAspAsnLeuHisLeuProIleLeuProLeuAsnGlyAspAsnGluLysArgLys 927
 QY 1789 TTTCTTGAATCTGAAGATGGGAGTAGACAGCAACAGCAACATATAGT----- 1836
 Db 928 ---GlnAsnSerGlnGluGlyAspGlnAlaHisProLysIleLysSerMetIleProGlu 946
 QY 1837 AGTGCCCAACATCGTCTGTATTTCCATAGTAGTACCTGCTTACCTCTCCGTCACAGAA 1896
 Db 947 SerGlySerSerSer-----HisThrGlu-----LysGlu 956
 QY 1897 ATGGAAGTAGATAGTGAAGATGAAGAGATCTCTGAATGG----- 1935
 Db 957 GluGluAsnGluGluGluGluLysProGluGlnHisLysGlnGluLysAspGln 976
 QY 1936 ---CTAAGAGAAAACCACTTACACAAATGAA 1965
 Db 977 GluLysArgGluLysValValAspAspMetGlu 987
 RESULT 10
 Y373 BOVIN
 ID Y373 BOVIN STANDARD; PRT; 1453 AA.
 AC Q9TU23;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein KIAA0373.
 GN KIAA0373.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Trachea;
 RA Jovov B., Ripoll P.J., Benos D.J.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC

| | | | |
|----|------|---|------|
| Db | 898 | LeuAsnAspLeuThrAsnGluLeuGlnAsnLysGlnLysAlaTyrGlyLysValLeuArg | 917 |
| Qy | 922 | GTGACTCGTCAGCAAGAAGAGAGCTTTAAATGGAATGATTATATGGAGAAACCAATGCAAAAT | 981 |
| Db | 918 | -----GluLysASPalaVal | 922 |
| Qy | 982 | ATTGATGTCATGAAGAGCTTCCAGCCAGAGAAAAAGAAATCGTAG- | 1039 |
| Db | 923 | AspGlnGluAsnAsnGluLeuLysArgGlnIleLysArgLeuThrSerGlyLeuGlnGly | 942 |
| Qy | 1030 | -----GATGGGAAAAACACATTTGTTCCAGCAAAATGCAGATTGTTGATAAAAAC | 1077 |

Db 943 LysProLeuIleAspAsnLysGlnSerLeuIleGluGluLeu-----Gln 957

| | | | | | | |
|------|----|---|------|----|---|------|
| 1078 | Qy | AGGCGCTTACAGCTTTTAGATGGGCAATATGAAGTAGCCATCGAGGAAATCGGAAGATGT | 1131 | Db | LylysileLysLysLeuGluSerGlnLeuGluAraGluValAspGluAraGluMetLys | 977 |
| 958 | Db | LylysileLysLysLeuGluSerGlnLeuGluAraGluValAspGluAraGluMetLys | 977 | Qy | CCATATAGCAAGAAAGAGCA | 1138 |
| 978 | Db | ProMetLysGlnLysSerAlaArgGluGluValIleArgTrpGluGluGlyLysLysTrp | 997 | Qy | GAGACTATTCTGATGGGAAGAGCGCTCCCTCCATTCGAAACATTTCTCAGGGACCTACG | 1224 |
| 1165 | Qy | GAGACTATTCTGATGGGAAGAGCGCTCCCTCCATTCGAAACATTTCTCAGGGACCTACG | 1224 | Db | GlnThrLysIleGluGlyIleArgAsnLysLeuLysGlu | 1015 |
| 998 | Db | GlnThrLysIleGluGlyIleArgAsnLysLeuLysGlu | 1015 | Qy | TTGCAGTTCCTCTTCGTTGGCAGGAGAGACCAATGATAAATCTACGGCTCCTATTGCC | 1284 |
| 1225 | Qy | TTGCAGTTCCTCTTCGTTGGCAGGAGAGACCAATGATAAATCTACGGCTCCTATTGCC | 1284 | Db | TyrIleLeuThr | 1016 |
| 1016 | Db | TyrIleLeuThr | 1016 | Qy | AAACCTCTTCCCACTAGAAATTCCAGAGAGTCTCCATCAGAGAAAACAACCGCTGGTTCAGTT | 1285 |
| 1285 | Qy | AAACCTCTTCCCACTAGAAATTCCAGAGAGTCTCCATCAGAGAAAACAACCGCTGGTTCAGTT | 1285 | Db | AspLeuPheAlaLysAlaAspLysGluLysLeuThrLeuGlnArgLys | 1043 |
| 1027 | Db | AspLeuPheAlaLysAlaAspLysGluLysLeuThrLeuGlnArgLys | 1043 | Qy | AAACCTACTCAAATATTGCTGTGTTAAGAATCATGTGCTACAGAT | 1345 |
| 1345 | Qy | AAACCTACTCAAATATTGCTGTGTTAAGAATCATGTGCTACAGAT | 1345 | Db | LysThrThr | 1044 |
| 1044 | Db | LysThrThr | 1044 | Qy | CTACAAACAAGAAAAGAAAGATCTCCAAATGAAACCCACAAAATTAAGA | 1443 |
| 1390 | Qy | CTACAAACAAGAAAAGAAAGATCTCCAAATGAAACCCACAAAATTAAGA | 1443 | Db | ArgValLeuGluSerGluLysGluLeuGluLysLysArgAsnLeuAspLeuGlu | 1076 |
| 1057 | Db | ArgValLeuGluSerGluLysGluLeuGluLysLysArgAsnLeuAspLeuGlu | 1076 | Qy | ATATTATTACAGTTTCTATATACACAATACAAAGCAACAACCTGAAGCAAGAGATGAC | 1503 |
| 1444 | Qy | ATATTATTACAGTTTCTATATACACAATACAAAGCAACAACCTGAAGCAAGAGATGAC | 1503 | Db | AsnAspIleSerTyrMetArgSerHisGlnAlaLeuProArgAspSerValIleGluAsp | 1096 |
| 1077 | Db | AsnAspIleSerTyrMetArgSerHisGlnAlaLeuProArgAspSerValIleGluAsp | 1096 | Qy | CTCATGTGCCCTTGGTGTACTCTGAACTGCGCGAAACTTTATAGTTTACTCAAGCATCTT | 1563 |
| 1504 | Qy | CTCATGTGCCCTTGGTGTACTCTGAACTGCGCGAAACTTTATAGTTTACTCAAGCATCTT | 1563 | Db | LeuHisLeuGlnAsnLysTyrLeuGln | 1115 |
| 1097 | Db | LeuHisLeuGlnAsnLysTyrLeuGln | 1115 | Qy | AAACTCTGCCATAGCAGATTATCTTCAACTATGTTTATCATCCAAAGAGTGCTAGG | 1620 |
| 1564 | Qy | AAACTCTGCCATAGCAGATTATCTTCAACTATGTTTATCATCCAAAGAGTGCTAGG | 1620 | Db | SerLysAspAlaTyrSerArg | 1116 |
| 1116 | Db | SerLysAspAlaTyrSerArg | 1116 | Qy | ATAGATGTTTCTATCAATGAGTGTATTGATGGCTCTCTATCAGGAAATCTCAGGATATT | 1680 |
| 1621 | Qy | ATAGATGTTTCTATCAATGAGTGTATTGATGGCTCTCTATCAGGAAATCTCAGGATATT | 1680 | Db | IleAsp | 1128 |
| 1128 | Db | IleAsp | 1128 | Qy | CATCGCAACCTGGATTGCTTTTACTGCGCAACGACCGATTAAAG | 1737 |
| 1681 | Qy | CATCGCAACCTGGATTGCTTTTACTGCGCAACGACCGATTAAAG | 1737 | Db | GlnArgGluAsnLeuLysLeuSerSerGluAsnIleGluLeuLysPheGlnLeuGluGln | 1160 |
| 1141 | Db | GlnArgGluAsnLeuLysLeuSerSerGluAsnIleGluLeuLysPheGlnLeuGluGln | 1160 | Qy | ACACATATTCTTGTCGACGGCCAAA | 1788 |
| 1738 | Qy | ACACATATTCTTGTCGACGGCCAAA | 1788 | Db | AlaAsnLysAspLeuProArgLeuLysAsnGlnValArgAspLeuLysGluMetCysGlu | 1180 |
| 1161 | Db | AlaAsnLysAspLeuProArgLeuLysAsnGlnValArgAspLeuLysGluMetCysGlu | 1180 | Qy | TTTCTTGATCTCAAGATCGGGAAGTAGACACGAA | 1842 |
| 1789 | Qy | TTTCTTGATCTCAAGATCGGGAAGTAGACACGAA | 1842 | Db | PheLeuLysGlnLysGlyLysAlaGluValGluArgLysLeuGlyValArgGlySerGly | 1200 |
| 1181 | Db | PheLeuLysGlnLysGlyLysAlaGluValGluArgLysLeuGlyValArgGlySerGly | 1200 | Qy | CACAATCGTCTGTATTCCATAGTAGATACCTGCTTACCTCTCCGTCACCAAGAAATGGAA | 1902 |
| 1843 | Qy | CACAATCGTCTGTATTCCATAGTAGATACCTGCTTACCTCTCCGTCACCAAGAAATGGAA | 1902 | Db | ArgSer | 1201 |
| 1201 | Db | ArgSer | 1201 | | | |

```

Db      1201 ArgSer-----1202
      :
      :

```



```

Db      316 luLeuAsnSerIleArgGluLeuAsnThrAlaLys-----ValIleAlaA 331
QY      857 TTTTCAGTAATGAATTTGAACCTAGTAACAGCACCATATGGTGAAGCTTACTCGTGCTAT 916
Db      331 spAspSerLysGlnThrProGluAsnGluAspLeuLeuGlnLeuThrL 351
QY      917 TTAGAGTGAAGTCCAGGAGAGAGAGAGATTTAATGGAATGATTAAATGAGAAACCAATG 976
Db      351 ysGluLysLeuAlaGlnCysGluLysGluCysLeu---ArgLeuSerSerIleThrAspG 370
QY      977 AAAATATTGATCTCAATGAGAGCTTCAGCC----- 1008
Db      370 luAlaAspGluAspAsnGluAsnLeuSerAlaLysSerSerSerAspPheIlePheLeuL 390
QY      1009 -----AGAGAAACAGAAATCGTGAGATGGGAGAAAGACAT 1045
Db      390 ysLysGlnLeuIleLysGluArgArgThrLysGluHisLeuGlnAsnGlnIleGluThrP 410
QY      1046 TTGTTGCACAAATG-----ACAGTATTGATAAAMAAACAGGCGCTTACAGC 1090
Db      410 heIleValGluLeuGluHisLysValProIleIleAsnSerPheLysGluArgThrAspM 430
QY      1091 TTTTAGATGGGAATATGAATAGCAGATGAGCAATGGAAGATGTCCAATAAGCAAGA 1150
Db      430 erLeuGluAsnGluLeuAsnAsnAlaAlaLeuLeuGluHisThrSerAsnGluLysA 450
QY      1151 AAAGAGCAACATGGGAGACTATTTCTTGATGGGAAGAGGCTGCCTCCATTCGAAACATTTT 1210
Db      450 snAlaLysValLysGluLeuAsnAlaLysAsnGlnLysLeuValGluCysGluAsnAspL 470
QY      1211 CTCAGGACCTACG-----TTGAGTTCACTCTTCGTT 1243
Db      470 euGlnThrLeuThrLysGlnArgLeuAspLeuCysArgGlnIleGlnThrLeuLeuIleT 490
QY      1244 GGACAGGAGACCAATGATAATCTACGGTCTCTATTGCCAAA----- 1287
Db      490 hrAsnSerValSerAsnAsp---SerLysGlyProLeuArgLysGluGluIleGlnPheI 509
QY      1287 ----- 1287
Db      509 leGlnAsnIleMetGlnGluAspSerThrIleThrGluSerAspSerGlnLysValV 529
QY      1288 -----CCTCTTGCCACTAGAAATTC----- 1308
Db      529 alThrGluArgLeuValGluPheLysAsnIleIleGlnLeuGlnGlnLysAsnAlaGluL 549
QY      1309 -----GAGAGTCTCCATCAGGAAACCAAGC 1333
Db      549 euLeuLysValValArgAsnLeuAlaAspLysLeuGluSerLysGluLysSerLysG 569
QY      1334 CTGGTTTCAGTTAAA---CCTACTCAAACTATT---GCTGTTAAAGCAATCATCTACTACA- 1386
Db      569 lnSerLeuGlnLysIleGluSerGluThrValAsnGluAlaLysGluAlaIleIleThrL 589
QY      1387 -----GATCTACAAACAAGAAAGAAAG----- 1410
Db      589 euLysSerGluLysMetAspLeuGluSerArgIleGluGluLeuGlnLysGluLeuGluG 609
QY      1411 -----GATCTCAATGAATGAACCGCAAAATTAAGAAATATTTATCAGTTTC 1459
Db      609 luLeuLysThrSerValProAsnGluAspAlaSer----- 620
QY      1460 TCTATAACAACAATACA---AGGCAACAACTGAAGCAAGAGATGACCTGATGCCCTT 1516
Db      621 ---TyrSerAsnValThrIleLysGlnLeuThrLysArgAspLeuGluSerGlnV 640
QY      1517 GGTGACTCTGAAC-----TGCCGCAAACTTTATAGTTTACTCAAGC 1558
Db      640 alGlnAspLeuGlnThrArgIleSerGlnIleThrArgGluSerThrGluAsnMet-Ser 659
QY      1559 ATCTTAATCTGCGATAGAGATTATCTTCAACTATGTTTATCATCCAAAGCTGCTA 1618

```

```

Db      660 LeuLeuAsnLysGluIleGlnAspLeuTyrAspSerLysSerAspIleSerIleLysLeu 679
QY      1619 GGATAGATGTTTCTATCAATGAGTGTATGATGCTCTCTATGATGAGAAATCTCTCAGATA 1678
Db      680 Gly-----LysGluLysSerArgIleLeuAla-----GluGluArgPheLysLeu 695
QY      1679 TTCATGCCCAACCTGGATTGCTTTAGTCGCAACGACCAAGTAAAGAACACCTATCA 1738
Db      696 LeuSerAsnThrLeuAspLeuThrLysAlaGluAsnAspGlnLeuArgLysArgPheAsp 715
QY      1739 CACATATT-----CTTGTCGACGGCCAAACGAACAAACAAACAGCATGCTGTAATTT 1791
Db      716 TyrLeuGlnAsnThrIleLeuLysGlnAspSerLysThrHisGluThrLeuAsnGluLys 735
QY      1792 CTT-----GAATCT 1800
Db      736 ValSerCysLysSerLysLeuSerIleValGluThrGluLeuLeuAsnLeuLysGluGlu 755
QY      1801 GAGATGGGGAAGTAGAACAGCAACAAACATATAGTAGTGGCCACAAATCGTCTGTATTT 1860
Db      756 GlnLysLeuArgValHisLeuGluLysAsnLeuLysGlnGluLeuAsnLysLeuSerPro 775
QY      1861 CATAGTAGTACCTGCTTACCTCTCCGTCACAGAAATGGAAGTAGATAGTAGAGTAA 1920
Db      776 GluLysAspSerLeuArgIleMetValThrGlnLeuGlnThrLeuGlnLysGluArgGlu 795
QY      1921 AAGATCTCTGAATGGTAAAGAGAAACCAATTCACAAATTCGACAGATTTTCTGAT--- 1977
Db      796 AspLeuLeuGluGluThrArgLysSerCysGlnLysLysIleAspGluLeuGluAspAla 815
QY      1978 GTTAATGAGGAGAGAGAGAGAGTGAAGTGAAGTCTGGAATCTCCATCATGAGCATGG 2037
Db      816 LeuSerGluLeuLysLysGluThrSerGln-----LysAspHisHisIleLysGln--- 832
QY      2038 TTTATTGCTGACAAATCAATGAAATCATGCTGCTGATGCTGTTGTGAAATATATGACAG 2097
Db      833 LeuGluGluAspAsnAsnSerAsn-----IleGluThrPyrGlnAsn 846
QY      2098 AAAATA-----ATTGAAGAAATTTATGTCGAAACTTCATGCTTCATCTAGTCAGCATG 2151
Db      847 LysIleGluAlaLeuLysLys----- 853
QY      2152 CATGACTTAAATCTTTATTAGCATAATGTCAATAGATAAAAGCTGTACCAAGCTCCGTGAA 2211
Db      854 ---AspTyrGluSerVal---IleThrSerValAspSerLysGlnThrAspIleGluLys 871
QY      2212 ATCAGCAAAAATTAGAAAAGGGGGAATCTGCTCCCTGCCAAACGAAAGAAATTAAGTAA 2271
Db      872 LeuGlnTyrLysValLysSerLeuGlu-----LysGluIleGluGlu 885
QY      2272 GAACAAATGGGACACAAATGATTTAGTGAATTAATCAAAAGAGAAAGCTTTGGAA 2331
Db      886 AspLys-----IleArgLeuHisThrTyrAsnValMetAspGluThrIleAsn 901
QY      2332 ACAGATAGTCTCTCAGGGGTTTCAAAACAGACAGCAAAAACTCTGAAAAAGCTCTA 2391
Db      902 AspAspSer-----LeuArgLysGluLeuGluLysSerLysIle-----AsnLeu 916
QY      2392 ACCCATGTTATGGACAA 2409
Db      917 ThrAspAlaTyrSerGln 922

```

```

RESULT 12
UBP3 YEAST
ID_UBP3 YEAST STANDARD; PRT; 912 AA.
AC Q01477;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 3 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 3) (Ubiquitin-specific processing protease 3)
DE (Deubiquitinating enzyme 3).
GN UBQ3 OR YER151C.

```


QY 986 ATGTCAATGAAGAGCTTCACCGCAGAGAAACGAAATCGTGAGGATGGGAAAGACAT 1045
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 275 spAlaHisAspLysProHisAlaSerLysAspAlaHisGluLeuLysLysThrG 295
 QY 1046 TTGTTCACAAATGACGATTGTTGATAAAACAGCGGCTTACAGCTTTAGATGGGAAT 1105
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 295 luValLysLys----- 298
 QY 1106 ATGAAGTAGCCATGACGGAATGGAAGATGTCCTAATGAAGCAAGAAAGACACATGGG 1165
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 299 --GluAspAluLysGlnAspArgAsnGluLysValIleGlnProGlnAlaThrValL 318
 QY 1166 AGCATATTCTTGATGGGAAGAGCTGCGCTCCATTCGAAACATTTTCTCAGGAGCACTAGT 1225
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 318 euProValValAspLysLysGlu-----ProGluGluSerValGluGluAsn----- 333
 QY 1226 TCGAGTTCACCTCTGTTGGACAGAGACGACCAATGATAAATCTACGCTCTATTGCG- 1284
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 334 -----ThrSerLysThrSerSerProSerProSerProAlaA 347
 QY 1285 -----AAACCTCTTGCCATAGAAATTCAGAGAGTCTCCATCAGGAAACAAAGC 1333
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 347 laLysSerTrpSerAlaIleAlaSerAspAlaIleLysSerArgGlnAlaSerAsnLysT 367
 QY 1334 CTGGTTCAGTAAACCTACTCAACTATTGCTGTTAAAGATCATCTGACTACAGATCTAC 1393
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 367 hrValSerGlySerMetValThrLysThrProLysSerGlyThrAlaGlyValSers 387
 QY 1394 AAACAGAAAGAAAGGATAGTCCAAATGAAACCGCAAAATTAAGAATATTTTATC 1453
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 387 erThrAsnMetAlaAlaThrIleGlyLysSerSerProLys----- 402
 QY 1454 AGTTCTCTATACCAACATACAGGCAACAACTGAAGCAAGAGACCTGCATTCGCC 1513
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 403 -----LeuSerLysGlnProGlnLysLysAspLysLysTyrValP 416
 QY 1514 CTTGGTGTACTGTCACTCCCGCAACTTATAGTTTACTCAAGCATCTTAAACTCTGCC 1573
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 416 roProSerThrLysGlyIleGluProLeuGlySerIleAla-----LeuArgMetCysP 434
 QY 1574 ATAGCAGATTATCTTCAACTATGTTTATCATCCAAAGGTCTAGGATAGATGTTCTA 1633
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 434 heAspProAspPheIleSerTyrValLeuArgAsnLysAspValGluAsnLysIleProV 454
 QY 1634 TCAATGAGTGTATGATGCTCCTATGCGAAGAAATCCTCAGATATTCATGCCAACCTG 1693
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 454 alHisSerIleIle-----ProArgGlyIleIleAsnArgAlaA 467
 QY 1694 GATTGTCTTTAGTCGCAACGACCATGTTAAGAGAACCTATCACATATTTCTTGT 1753
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 467 snIleCysPhe-----MetSerSerValLeuGlnValLeuLeuTyrC 481
 QY 1754 CGAGGCCA-----AACGAAACAAAGCAAGCATG- 1782
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 481 ysLysProPheIleAspValIleAsnValLeuSerThrArgAsnThrAsnSerArgValG 501
 QY 1782 ----- 1782
 DB 501 lyThrSerSerCysLysLeuLeuAspAlaCysLeuThrMetTyrLysGlnPheAspLysG 521
 QY 1783 -----TCTGAATTTCTGAATCTGAAGATGGG-----GAAGTAGAACAGCAAA 1825
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 521 luThrTyrGluLysLysPheLeuGluAsnAlaAspAspAlaGluLysThrThrGluSerA 541
 QY 1826 GAACATATAGTGTGCCACATCGTCTGTATTTCATAGTATGATCTGCTTACCTCTCC 1885
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 541 spAlaLysLysSerLysSerLysSerPheGlnHisCysAlaThrAlaAspAlaValL 561
 QY 1886 GTCCACAAAGAAATGGAAGTAGATAGTGAAGATGAAAGGATCTGATGGCTTAAGAGAA 1945
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 561 ysProAspGluPhe-----TyrL 567
 QY 1946 AAACATTACAAATGAAGAGTTTCTGATGTTAATGAAGAGAGAAAGAAAGTATGA 2005

DB 567 ysthrLeuSerThrIleProLysPheLysAspLeuGlnTyrGlyHisGlnGluAspAlaG 587
 QY 2006 AACTCTGGAATCTCCATGTCATCAAGCAT-----GGGT 2038
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 587 luGluPheThrHisLeuLeuAspGlnLeuHisGluGluLeuIleSerAlaIleAspG 607
 QY 2039 TTATTGCTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 2098
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 607 lyLeuThrAspAsnGlu-----IleGlnAsnMetLeuGlns 619
 QY 2099 AAATAATTAAGAAGAAATTTATGCGAAACTTCATGCTTCTATGTCAGCATGCAT----- 2154
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 619 erIleAsnAspGluGlnLeu---LysValPhePheIleArgAsnLeuSerArgTyrGlyL 638
 QY 2155 -----GACTTTAATCTTATTAGCAATAATGTCATAGATAAAGCTGTTCACCAAGCTCGTG 2209
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 638 ysAlaGluPhe-----IleLysAsnAlaSerProArgLeuLysG 651
 QY 2210 AAATGACGACAAAATTAGAAAAGGGGAATCTCTTCCCTGCAAGCAAGAAATAACTG 2269
 DB : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 651 luLeuIleGluLysTyr-----GlyValIleAsnAspAspSerThrG 665
 QY 2270 AAGAACAAATGGACAGCAAAATGATGATTTAGTGAATTAATCACTCAAGAGAAAGCTTGG 2329
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 665 luGlu-----AsnGlyTyrHisGlu----- 671
 QY 2330 AAACAGATAGTGTCTCAGGCTTTCAAAACAGACGACCAAAACAAAAA 2376
 DB ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 DB 672 -----ValSerGlySerSerLysArgGlyLysLysThrLys 683
 RESULT 13
 PCPI SCHPO STANDARD; PRT; 1208 AA.
 AC Q92351;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Spindle pole body protein pcpi.
 GN PCPI OR SPAC69.06C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 [1]
 RN SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RP MEDLINE=21852775; PubMed=11864908;
 RA Flory M.R., Morshew M., Joseph J.D., Means A.R., Davis T.N.;
 RT "Pcplp, a Spcl10p-related calmodulin target at the centrosome of the
 RL fission yeast Schizosaccharomyces pombe.";
 RL Cell Growth Differ. 13:47-58(2002).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=972;
 RA MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourse J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford J., Rutter S., Saunders R., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of *Schizosaccharomyces pombe*,"
 RL Nature 415:871-880 (2002).
 CC -!- FUNCTION: Spindle pole body component that binds calmodulin.
 CC Overexpression of pcpi causes the formation of supernumerary SPB-
 CC like structures and disrupts both mitotic spindle assembly and
 CC chromosome segregation.
 CC -!- SUBCELLULAR LOCATION: Spindle pole body.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z81317; CAB03608.1; -.
 DR EMBL; AF348506; AAK31344.1; -.
 DR PIR; T39068; T39068.
 DR GeneDB SPombe; SPAC69.06c; -.
 KW Calmodulin-binding; Coiled coil.
 FT DOMAIN 151 375 COILED COIL (POTENTIAL).
 FT DOMAIN 387 803 COILED COIL (POTENTIAL).
 FT DOMAIN 874 1091 COILED COIL (POTENTIAL).
 FT DOMAIN 1177 1204 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADD42424 CRC64;

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 0.00236 | Length: | 1208 |
| Score: | 157.00 | Matches: | 184 |
| Percent Similarity: | 35.51% | Conservative: | 153 |
| Best Local Similarity: | 19.39% | Mismatches: | 300 |
| Query Match: | 2.04% | Indels: | 312 |
| DBs: | 1 | Gaps: | 40 |

US-09-874-162a-7 (1-4409) x PCPI_SCHPO (1-1208)

| | | | |
|----|-----|--|-----|
| QY | 127 | CTGCCGACCTCATGACACATCGAGGACCAACACATCGATACAGATCCACGGTTTAA | 186 |
| DB | 127 | LeuThrAspLeuSerLysAsnSerGluLeuAspHisThrAspGly | 141 |
| QY | 187 | GAATAACAGATACACGACCAACCTATGTTCCCTGAGTACATAAATAGATTTCATG | 246 |
| DB | 142 | -----GluLeuProAlaAsnAlaAlaLeuThrLeu----- | 151 |
| QY | 247 | ACAGATGTCGCCGCGAGACGAGTCCCTAAGAGAGATTCAGCCGAGCTCTCG | 306 |
| DB | 152 | -----ArgGluGlnGluLysValLeuGluLysVal-SerArgGluAsnPh | 166 |
| QY | 307 | CTGACTC--TGTCCAGCTAGTCTCTGAGGGAATGTGCCACTCCCGACGCCACAGC | 363 |
| DB | 166 | eGlyLeuArgIleLysIleValCysLeuGlu-LysArgLeuGluSerMetAlaProGluG | 186 |
| QY | 364 | AGTGGAGCCTTACTCCCGCGTACGCCACCCATCCCTCTCTTCCATCCGCGAGC | 423 |
| DB | 186 | inLeLysGlu--AlaValLysAsnVal-----GluL | 197 |
| QY | 424 | AGCACTCCGACAGACCAACAGATCTATAGATTCTTGAATCGGAATCTCATGCA | 483 |
| DB | 197 | euHisAlaGluArgAlaAsnLeuGlnLeu-GlnLeuLysArgThrGluSerLeuLeu--- | 215 |
| QY | 484 | CCAATATTTTGGACAGAACTTTACTTACATGCTCATCGAACTCCAGAACAAACATC | 543 |
| DB | 216 | -----GlnLysSer | 218 |
| QY | 544 | AAAGGAAACATTAAAGTTGAT-----GATATGTTATCAAAAGTAGAGAAATG | 594 |

| | | | |
|----|------|---|------|
| DB | 219 | GluAspLysAsnPhelLysLeuGluGluLysValAspTyrLeuSerLysValAsnAspVal | 238 |
| QY | 595 | AAAGGAGACAGAAATCTCATAGCTTGTGACGCTCATTTGACAGCTTACGTTTACTGTTTC | 654 |
| DB | 239 | GluGlnSerGlnAsnValLysValPheThrGluArgIleArg----- | 252 |
| QY | 655 | TTCACAAAAAT-----GATAGCCATCCACAACTCAGAAAT | 693 |
| DB | 253 | PheLeuGluAsnAlaLeuGluLysValGlnArgGluLysAspSerLeuSerThrGluMet | 272 |
| QY | 694 | GAACAAATTTCTGTACCTGGAAGTCTGCTGTGAAAGTTTGCACAAAAAGAAAG | 753 |
| DB | 273 | GluGluAspLysSerAsnLysGluVal----- | 281 |
| QY | 754 | GATGTAAGTTGTCCTAATAAGCAAGTTCCACAGGT-----AAAAAGCAGGTG | 801 |
| DB | 282 | AspTyrGluTyrGluLeuArgGlnLeuGlnAsnArgLeuAspGluLeuSerGluGluLeu | 301 |
| QY | 802 | CCTTTGATTCCTGACCTCAATCAACAAACCCGGAATTTCCCTCCCTTCAGCTTCC | 861 |
| DB | 302 | AspValAlaGlnAspLeuThrGluLysGluAspGluLeuAlaThrLeuLysArgGln | 321 |
| QY | 862 | AGTAATGAATTTGAACCTAGTAACAGC----- | 888 |
| DB | 322 | IleGluGluLysGluAsnSerSerAlaPheGluAsnGluGluAsnSerTyrVal | 341 |
| QY | 889 | CATATGGTCAAGCTCTTACTCGTTGCTATTTAGAGTCTGCTCCAGGAGAGAGAGTTT | 948 |
| DB | 342 | HisLeuGlnGluAspTyrAlaIleLeu----- | 350 |
| QY | 949 | AATGGAATGATTAAATGGAGAACCAATGAAATATTTGATGTCATGAAAGAGCTTCCAGCC | 1008 |
| DB | 351 | -----GlnAlaLysCysAspGluPheAlaAspArgIleGlnValLeuThrAla | 366 |
| QY | 1009 | AGAAGAAACGAAATCGTGAG----- | 1029 |
| DB | 367 | AspLeuGluLysGluLysGluAsnGlnIleMetHisGluSerGluAlaSerIleGlyLeu | 386 |
| QY | 1030 | ---GATGGGAAAGACATTTGTCACAAATGACAGTATTTGATAAAACAGGGCGTTA | 1086 |
| DB | 387 | ThrAspSerMetGlnValHisThrLeuGlnGluGlnHisLysAlaAsnGluLeu | 406 |
| QY | 1087 | CAGCTTTTAGTGGGAATAT-----GAAGTA | 1113 |
| DB | 407 | GluPheLeuHisAspGlnIleSerArgMetAsnGluGluGlyLysAsnPheGluAspIle | 426 |
| QY | 1114 | GCCATGCCAG-----GAAATGGAAGAATGTCCTCAATAAGCAAGAAAGACACACA | 1161 |
| DB | 427 | MetLeuGlnPheArgSerLeuGluGluGluArgAspValLeuSerLysLeuGlnThr | 446 |
| QY | 1162 | TGGGAGACTATTCTTGATGGGAGAGGCTGCTCCATTCGAAACATTTTCTCAGGAGCT | 1221 |
| DB | 447 | LeuGluAspAspAsnSerLeuArgLeuMetThrSerSerLeuGlyAsnGlnIleGlu | 466 |
| QY | 1222 | ACGTTGTCAGTTCCTCTTCGTTGGACAGGAG----- | 1254 |
| DB | 467 | SerLeuArgThrGlnAsnArgGluLeuAspGluLysAsnHisLeuArgLeuLeuAla | 486 |
| QY | 1255 | -----ACCAATGATAAATCTACGGCT-----CCTATTGCCAAA | 1287 |
| DB | 487 | SerLysAsnSerAspLysAlaLeuAlaGluThrAsnIleArgLeuGlnGluValThrLys | 506 |
| QY | 1288 | CCTCTGCCACT-----AGAAATTCAGAGAGTCTCCAT----- | 1320 |
| DB | 507 | GluLeuGluThrLeuArgMetLysAsnSerAsnAspLeuAsnGluIleHisAspLeuArg | 526 |
| QY | 1321 | CAGGAAACAAAGCCTGGTTTCAGTTAAACCTACTCAAACTATTGCTGTTAAAGAAATCTTG | 1380 |
| DB | 527 | GluGluAsnGluGlyLeuThrLeuLysIleAspSerIleThrLysGluLysArgLeu | 546 |
| QY | 1381 | ACTACAGATCTCAACAAAGA-----AAAGAAAGGAT | 1413 |

547 IleAsnGluLeuGluGlnArgIleLysSerTyrGluValAsnValSerGluLeuAsnGly 566
1414 ACTCAAATGAAACCGACAAATTAAGA-----ATATTT 1449
567 ThrIleAspGluTyrArgAsnLysLeuLysAspLysGluGluThrTyrAsnGluValMet 586
1450 TATCAGTTCTCTAT---AACAAACAAATCAAGCAACAACTGAAGCAAGATCACCTG 1506
587 AsnAlaPheGlnTyrLysAspAsnAspLeuArgArgPheHisGluSerIleAsnLysLeu 606
1507 CATTCGCCCTTGCTGACTCTGAACGCGCAACCTTATAGTTTACTCAAGCATCTTAA 1566
607 GlnAsp-----ArgGluLysGluLeuThrSerAsnLeuGluLys-LysAs 621
1567 CTCTGCATAGCAGATTTATCTTCAACTATGTTTATCATCCAAAAGGTGCTAGGATAGAT 1626
621 neuValIleSerSerLeuArgGluThrValAlaMetLeuGluLys----- 636
1627 GTTCTCATCAAGTAGTGTATGATGCTCTCTATGAGCAAAATCTTCAGATATTCATCGC 1686
637 -----GluArgGluSerIleLysLysTyrLeuSe 646
1687 C-----AACCTGGATTGCTTTAGTCGCAACGACCCAG----- 1720
646 rGlyAsnAlaLysAspLeuAspAsnThrAsnLeuMetGluIleLeuAsnAspLysIleSe 666
1721 -----TTAAGAGAACCTATCACACATA----- 1744
666 rValLeuGlnArgGlnLeuThrAspValLysAspGluLeuAspValSerGluGluLys 686
1745 -----TTCCTGTGTGAGGCGCAAAACGAAACGAAAGCAGCATCTCTGAATTTCTTGA 1797
686 gGluGluAlaIleValAlaGlyGln--LysLeuSerAlaSerPhe--GluLeuMetSer 704
1798 TCTGAAGATGGGAAGTAGACAGACAGAACATATAGTAGGCGCACAACTGCTGAT 1857
705 AsnGluLysGlnAlaLeuGluLys-----TyrSerSerLeuLysAsnGluLeu-- 721
1858 TTCCATAGTAGTACTGCTTACTCTCCGTCACAAAGAAAGGAAGTAGATGAAGAT 1917
722 IleAsnAlaGlnAsnLeuLeuAspArgArgGluGluGluLeuSerGluLeuSerLys 741
1918 GAAAGGATCCTGAATGGCTAGAGAAACCAATACACAAATGAAGATTTCTCAT 1977
742 LeuPheGluGluArgLysIleArgSerGlySerAsnAspIleGluLysAsnLysGlu 761
1978 GTTAATGAAGAGAGAAAGATGATG---AAACTCTGGAATCTC---CATGTCATGAAG 2031
762 IleAsnValLeuAsnSerGluLeuAlaAspLysLeuAlaGlnIleArgHisLeuGluSer 781
2032 CATGGTTTATCTGACAAATCAATGAATCATGCTGTATGCTGTTGTAGAA---AAT 2088
782 AspLysMetGluLeuAspLysLeuValHisHisLeuAsnArgGlyIleGluAlaAsn 801
2089 TATGCACAGAAATATTAAGAGAAATTTATGTCGAAACTTCATGCTCATCTAGTCAGC 2148
802 IleGluGluAsnAlaValLysLysArgLeuCys-----LeuLeuMet 815
2149 ATGCATAGCTTAATCTTATTAGCAATAGTCAATA----- 2184
816 GlyCysAspTyrSerSerValSerIleLeuGlnIleValSerGlnIleGluHisPheVal 835
2185 GATAAGCTGTACCAAGCTCGTGAATGCGACGAAATTAAGAA----- 2229
836 AsnGlnGlnIleGlnThrIleArgSerLeuLysGlnGluLeuArgHisAspPheValGln 855
2229 ----- 2229
856 PheSerGlyLysLysGluGlnGluLeuSerA-SerPheGluLysPheGlyLeuGlyThr 875
2230 -----AAGGGGAATCTCTCCCTGCAACGAGAAATTAACTCAGACAAAT--- 2280
876 GluThrLysHisAspIleLeuAlaGlnArgAsnArgAsnValSerGluLysMetAsnAsp 895

2281 -----GGGACAGCAAAATGGATT--- 2298
896 LeuGluAsnAlaAlaGlnLysPhePheSerSerProAspArgLysAsnGlyTyrLeuTyr 915
2299 -----AGTGAATTAACATAAAGAGAAAGCTTTGGAA----- 2331
916 ProSerGluHisThrSerLysIleGluTyrLeuGluLysThrIleGluAspLeuLysLeu 935
2332 -----ACAGATAGTCTCTCAGGGTTTCA 2355
936 AlaLeuGlnAspGluLeuLysAsnArgAsnLeuMetAspLysSerSerTyrAsn 955
2356 AAACAGAGCAAAAAA 2370
956 LysGlnThrThrLys 960
RESULT 14
YCZ7_YEAST
ID YCZ7_YEAST STANDARD; PRT; 1755 AA.
AC P47098; P87194;
01-FEB-1996 (Rel. 33, Created)
30-MAY-2000 (Rel. 39, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposon Ty1 protein B.
GN TY1B OR YJ027W OR J1360.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=96109930; PubMed=8619316;
RA Zagulski M., Babinak B., Gromadka R., Migdalski A., Rytko J.,
Sulicka J., Herbert C.J.;
RT "The sequence of 24.3 kb from chromosome X reveals five complete open
reading frames", all of which correspond to new genes, and a tandem
insertion of a Ty1 transposon."
RL Yeast 11:1179-1186(1995).
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: Z49526; CAA89553.1; --
DR MEROPS; A11.003; --
DR SGD; S0003788; YJ027W.
DR InterPro; IPR001969; Asprotease AS.
DR InterPro; IPR001042; Pept_A113_Ty1.
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF01021; TYA; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
FT ACT SITE 461 461
FT NP_BIND 1631 1638 ATP (POTENTIAL).
SQ SEQUENCE 1755 AA; 198615 MW; 04E248A77FD7596F CRC64;

Alignment Scores:
Pred. No.: 0.00315 Length: 1755
Score: 155.50 Matches: 159
Percent Similarity: 36.70% Conservative: 99
Best Local Similarity: 22.62% Mismatches: 232
Query Match: 2.02% Indels: 213
DB: 1 Gaps: 44

US-09-874-162A-7 (1-4409) x YCZ7_YEAST (1-1755)


```
SPFQ_HUMAN
ID   SPFQ_HUMAN STANDARD; PRT; 707 AA.
AC   F23246; P30808;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-
DE   binding protein-associated splicing factor) (PTB-associated splicing
DE   factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
GN   SPFQ OR PSF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]_TaxID=9606;
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC   TISSUE=Fetal brain;
RX   MEDLINE=93194059; PubMed=8449401;
RA   Patton J.G., Porto E.B., Galceran J., Tempst P., Nadal-Ginard B.;
RT   "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT   factor.";
RL   Genes Dev. 7:393-406(1993).
RN   [2]
RP   SEQUENCE OF 312-707 FROM N.A.
RC   TISSUE=Fetal skeletal muscle;
RX   MEDLINE=90091812; PubMed=2480877;
RA   Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
RT   "Cloning and characterization of a myoblast cell surface antigen
RT   defined by 24.1D5 monoclonal antibody.";
RL   Development 105:723-731(1989).
RN   [3]
RP   SEQUENCE OF 48-68 AND 213-246.
RX   MEDLINE=93176127; PubMed=8439294;
RA   Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
RT   "Purification and characterization of a DNA-binding heterodimer of 52
RT   and 100 kDa from HeLa cells.";
RL   Biochem. J. 290:267-272(1993).
CC   -!- FUNCTION: Essential pre-mRNA splicing factor required early in
CC   spliceosome formation. Binds to the mammalian polypyrimidine
CC   tracts. Forms a complex with the polypyrimidine tract-binding
CC   protein (PTB). Seems to also bind DNA.
CC   -!- SUBUNIT: Heterotrimer of two 52 kDa and two 100 kDa subunits.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Comment=Additional isoforms seem to exist;
CC   Name=Long;
CC   IsoID=P23246-1; Sequence=Displayed;
CC   Name=Short;
CC   IsoID=P23246-2; Sequence=VSP_005855;
CC   -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
CC   -!- CAUTION: Was originally (Ref.2) thought to be myoblast cell
CC   surface antigen 24.1D5 and a possible membrane-bound protein
CC   ectokinase.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL; X70944; CAA50283.1; -.
DR   EMBL; X16850; CAA34747.1; -.
DR   PIR; A46302; A46302.
DR   PIR; S29770; S29770.
DR   HSSP; P11940; 1CVJ.
DR   SWISS-2DPAGE; P23246; HUMAN.
DR   Genew; HGNC:10774; SPFQ.
DR   GK; P23246; -.
DR   MIM; 605199; -.
GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.
```

```
DR   GO; GO:0006397; P:RNA processing; TAS.
DR   GO; GO:0006371; P:mRNA splicing; TAS.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 2.
DR   SMART; SM00360; RRM; 2.
DR   PROSITE; PS00102; RRM; 2.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
KW   Alternative splicing.
FT   DOMAIN      297..369   RNA-BINDING (RRM) 1.
FT   DOMAIN      371..452   RNA-BINDING (RRM) 2.
FT   REPEAT      9..27     3 X 3 AA REPEATS OF R-G-G.
FT   REPEAT      19..21     1.
FT   REPEAT      25..27     2.
FT   REPEAT      25..27     3.
FT   DOMAIN      10..266   GLN/GLU/PRO-RICH.
FT   DOMAIN      10..15    POLY-GLY.
FT   DOMAIN      20..27    POLY-GLY.
FT   DOMAIN      56..65    POLY-PRO.
FT   DOMAIN      67..71    POLY-GLN.
FT   DOMAIN      95..98    POLY-GLN.
FT   DOMAIN      99..103   POLY-PRO.
FT   DOMAIN      184..188   POLY-PRO.
FT   DOMAIN      571..574   POLY-ARG.
FT   DOMAIN      613..616   POLY-GLY.
FT   DOMAIN      635..641   POLY-GLY.
FT   VARSPLIC    663..707
FT   FTIDVSP     005855
FT   KPRF -> VRMIDVG (in isoform Short).
FT   /FTIDVSP     005855
FT   G -> R (IN REP. 3).
FT   CONFLICT    243..243
FT   SEQUENCE    707 AA; 76149 MW; 6DBD5EA95E235847 CRC64;
Alignment Scores:
Pred. No.:      0.00422      Length:      707
Score:          153.00       Matches:    120
Percent Similarity: 29.95%   Conservative: 60
Best Local Similarity: 19.97% Mismatches:    207
Query Match:      1.99%     Indels:      214
DB:               1        Gaps:         22
US-09-874-162A-7 (1-4409) x SPFQ_HUMAN (1-707)
Qy   2   CQCCCGCGGCTCGCAGCGCGACACGAGGGGGCTCTCGATGATGACCATGACAGGCA 61
Db   35  ProProGlyMetGlyLeuAsnGlnAsnArgGlyProMetGlyProGlyProGlyGlnSer 54
Qy   62  TCGCGCGCGCTCTCTTCTTCCAATACCTTCGCGGATTCGGGGGCTGGGACTCCACTTCC 121
Db   55  GlyProLysProProLleProProProPro-ProHisGlnGlnGlnGlnProProPr 74
Qy   122  CCACCTTGGCGGACTTCATCGACACATCGAGGACACACATCGATGATGATGATGATGATGAT 181
Db   74  oGlnGlnProProGlnGlnPro-----ProHisGln----- 86
Qy   182  TTTTAGAAAAACAAGATTACAGCAGCAACCTATGTCCTGCTGATGATGATGATGATGATGAT 241
Db   87  -----ProProHisProGlnPr 93
Qy   242  TCATGACAGATGTGCGCGCGAGAGAGAGAGTCCCTTAAAGAAGAGATTCAGCCGAGAC 301
Db   93  oHisGlnGlnGlnGlnPro-----ProProGlnAspSerSerLysPr 109
Qy   302  TCTCGTGTACTCTGTCCAGCTCAGTGCTCGAGGAATGTCCTCACTCCCGCCAGCCACA 361
Db   109  oVal-ValAlaGlnGlnGlyProAlaProGlyValGlySerAlaProProAlaSerS 129
Qy   362  GCAGTGAAGCTTACTCCCGCGTGACCCCGCCATCACC-----CCCTCTCTTCAT 415
Db   129  erSer-----AlaProProAlaThrProThrSerGlyAlaProProGlySerG 146
Qy   416  TCCGACGAGCACTCCGACAGCCACACAGATCTATAGATTTCTTCGAACTCGGATC 475
Db   146  lyProGlyProThrProThrProProProAlaVal----- 157
```

QY 476 TCATAGCACCATAATTTTGCACAGAACTCTTACTTACATGTCTCATCGAAACTCCAGAA 535
 Db 157 ----- 157
 QY 536 CAAACATCAAAAGGAAACATTTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGA 595
 Db 157 ----- 157
 QY 596 AAGGAGACAGAATCTCATAGCTTGTGAGCTCATTTGCAGCTTACGTTTACTGTTTCT 655
 Db 158 -----ThrSerAlap 161
 QY 656 TCCACAAAATGATAAGCATCACCAAACTCAGAAAATGAAACAAAATTTCTGTACCCTGG 715
 Db 161 roProGlyAlap roProThrProProSerSerGly----- 173
 QY 716 AAGTCCTGTTGTGAAAGTTGSCCAAAAAGAAAGGATGTAAGTTGTCCAAATAAGGC 775
 Db 174 -----ValProThrThrProProGlnAlaglyGlyProProProProAlaa 190
 QY 776 AAGTTCCC-----ACAGTAAAGACAGTGCCT----- 804
 Db 190 laValProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyGlyG 210
 QY 805 --TTGATTCCTGACCTCAATCAAAACAAACCCGGAATTTCCCGTCCCTTCAGTTTCCA 862
 Db 210 lyLysMetPro-----GlyGlyProLysProGlyGlyGlyProGlyLysSerThrProG 228
 QY 863 GTAATGAATTTGAACCTAGTAACCCATATGGTGAAGCTTACTCGTTGCTATTAGAG 922
 Db 228 lyGlyHisProLysProProHisargGlyGlyGlyGlu----- 240
 QY 923 TGACTCGTCCAGGAAGAAGAGAGATT----- 948
 Db 241 --ProArgGlyGlyArgGlnHisHisProProTyHisGlnGlnHisGlnGlyProp 260
 QY 949 --AATGGAATGATTAATGGAGAAACCAATGAATAATATTGATGTCAATGAAGAGCTTCCAG 1006
 Db 260 roProGlyGlyProGlyGlyArgSerGluGlyLysSerAspSerGluGlyPhelysA 280
 QY 1007 CCAGAAAGAAACGAAATCGTGAGGATGGGAAAGACATTTGTCACAAATGACACTAT 1066
 Db 280 laAsnLeuSerLeuLeuArgProGlyGlyLysThrTyThrGlnArgCysArgLeup 300
 QY 1067 TTGATAAAACAGCGCTTACAGCTTTTATAGTGGGAATAT----- 1107
 Db 300 heValGlyAsnLeuProAlaAspIleThrGluAspGluPhelysArgLeuPheAlalyst 320
 QY 1107 ----- 1107
 Db 320 yrGlyGluProGlyGluValPheIleAsnLysGlyLysGlyPheGlyPheIleLysLeug 340
 QY 1108 -----GAAGTAGCCATCGAGAAATCGAAGAAATGTCCAATAGCAAGA 1150
 Db 340 luSerArgAlaLeuAlaGluIleAlaLysAlaGluLeuAspAspThrProMetArgGlyA 360
 QY 1151 AAAGA-----GCAACATGGGAGACTATTCTTGATGGGAAGAGCGTGCCTC 1195
 Db 360 rgGlnLeuArgValargPheAlaThrHisAlaAlaLeuSerValArgAsnLeuSerP 380
 QY 1196 CATTC-----GAAACATTTTCTCAG-----GGACCTACGTTGCAGT 1231
 Db 380 roTyValSerAsnGluLeuLeuGluAlaPheSerGlnPheGlyProIleGluArgA 400
 QY 1232 TCATCTTT-----CGTGCACAGGAGAG-----ACCAATGATAAATCTA 1270
 Db 400 laValValIleValAspArgGlyArgSerThrGlyLysGlyIleValIleGluPheAlas 420
 QY 1271 CGGCTCCTATTGCCAAACCTCTGCCACTGCAAAATTCAGAGAGT-----CTCCATC 1321
 Db 420 erLysProAlaAlaArgLysAlaPheGluArgCysSerGluGlyValPheLeuLeuThrT 440

QY 1322 AGGAAAACAGCGCTTTCAGTTTAACTTACTCAAACTATTGCTGTTAAAGAAATCATTTGA 1381
 Db 440 hrThrProArgProValIleValGluProLeuGluGlnLeuAspGluAspGlyLeup 460
 QY 1382 CTACAGATCTACAAAACAAGA-----AAAGAAAAGGATACTCCAAAATGAAA 1426
 Db 460 roGluLysLeuAlaGlnLysAsnProMetTyxGlnLysGluArgGluThrProProArgP 480
 QY 1427 ACCGACAAAATTAAGAATATTTTATCAGTTTCTC-----T 1462
 Db 480 heAlaGlnHisGlyThrPheGluTyxGluTyxSerGlnArgTrpLysSerLeuAspGluM 500
 QY 1463 ATAACAACAATACAAGGCAACAACCTGAA-----GCAAGAGATGACCTG 1506
 Db 500 etGluLysGlnGlnArgGluGlnValGluLysAsnMetLysAspAlaLysAspLysLeu 519

Search completed: August 25, 2004, 18:44:25
 Job time : 154 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 25, 2004, 18:24:39 ; Search time 445 Seconds
(without alignments)
6252.224 Million cell updates/sec

Title: US-09-874-162A-7

Perfect score: 7690

Sequence: 1 cccgcgcgcgcgcgcag.....tacataaagtccaataatat 4409

Scoring table: BLOSUMP62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO.spool.p/US09874162/runat.25082004.171844.2180/app.query.fasta_1.4551
-DB=SPTREMBL_25 -QWTF=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosump62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09874162 @CGN 1 1 717 @runat.25082004.171844.2180 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DST TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvivirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 244.5 | 3.2 | 626 | 10 Q84UX5 | Q84ux5 zea mays (m |

| ID | Q84UX5 | PRELIMINARY; | PRT; | 626 AA. |
|----|--|--------------|------|---------|
| AC | Q84UX5; | | | |
| DT | 01-JUN-2003 (Tremblrel. 24, Created) | | | |
| DT | 01-JUN-2003 (Tremblrel. 24, Last sequence update) | | | |
| DT | 01-OCT-2003 (Tremblrel. 25, Last annotation update) | | | |
| DE | VEF family protein. | | | |
| GN | VEF101. | | | |
| OS | Zea mays (Maize). | | | |
| OC | Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; | | | |
| OC | Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; | | | |
| OC | PACCAD clade; Panicoideae; Andropogoneae; Zea. | | | |
| OX | NCBI_TaxID=4577; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=cv. B73; | | | |
| RA | Chandler V.L., Kaeppler S.M., Kaeppler H.F., Cone K.C.; | | | |
| RT | "Sequences from the Plant Chromatin Consortium (NSF Plant Genome | | | |
| RT | program Grant 9575930)."; | | | |
| RL | Submitted (APR-2003) to the EMBL/GenBank/DBJ databases. | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=cv. B73; | | | |
| RA | Bergstrom D.E., Springer N.M., Schmitt L., Guthrie E., Sidorenko L., | | | |

ALIGNMENTS

RESULT 1

| ID | Q84UX5 | PRELIMINARY; | PRT; | 626 AA. |
|----|--|--------------|------|---------|
| AC | Q84UX5; | | | |
| DT | 01-JUN-2003 (Tremblrel. 24, Created) | | | |
| DT | 01-JUN-2003 (Tremblrel. 24, Last sequence update) | | | |
| DT | 01-OCT-2003 (Tremblrel. 25, Last annotation update) | | | |
| DE | VEF family protein. | | | |
| GN | VEF101. | | | |
| OS | Zea mays (Maize). | | | |
| OC | Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; | | | |
| OC | Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; | | | |
| OC | PACCAD clade; Panicoideae; Andropogoneae; Zea. | | | |
| OX | NCBI_TaxID=4577; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=cv. B73; | | | |
| RA | Chandler V.L., Kaeppler S.M., Kaeppler H.F., Cone K.C.; | | | |
| RT | "Sequences from the Plant Chromatin Consortium (NSF Plant Genome | | | |
| RT | program Grant 9575930)."; | | | |
| RL | Submitted (APR-2003) to the EMBL/GenBank/DBJ databases. | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=cv. B73; | | | |
| RA | Bergstrom D.E., Springer N.M., Schmitt L., Guthrie E., Sidorenko L., | | | |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 244.5 | 3.2 | 626 | 10 Q84UX5 | Q84ux5 zea mays (m |

RA Sellinger D., Kaepler S.M., Cone K.C.;
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY232824; A084022.1; -
 DR InterPro; IPR007087; Znf_C2H2.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 SQ SEQUENCE 626 AA; 71401 MW; 604C0D46BDSAB8718 CRC64;

Alignment Scores:
 Pred. No.: 1,458-10 Length: 626
 Score: 244.50 Matches: 123
 Percent Similarity: 34.21% Conservative: 110
 Best Local Similarity: 18.06% Mismatches: 251
 Query Match: 3.18% Indels: 197
 DB: 10 Gaps: 22

US-09-874-162a-7 (1-4409) x Q84UX5 (1-626)

```

QY 436 GAGCCAAACACAGATCTATAGATTTCTCGAACTCGAATCTCATAGACCAATATTTTGG 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53 LysProValGluLeuTyAsnIleIleGlnArgAlaMetLysAsnProLeuPheIle 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 496 CACAGAACTCTTACTTACATCTCTCATCGAACTCCAGAACAAACATCAAGAGAAACA 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73 GlnArgCysLeuLeuTyAsnIleHisAlaArgArgLysLysArgIleGln----- 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 556 TTTAAAGTTGATGATATGTTATCAAGATGAGAGAAATGAAAGGAGAGCAAGATCTCAT 615
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 ----- 89
QY 616 AGCTGTGAGCTCATTTGCGAGTTAGTTACTGTTTCTTCTCCACAAAATGATAAGCCA 675
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 -----IleThrIleSerLeuSerGly----- 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 676 TCACCAAACTCAGAAAATCAACAAAATTCGTGTT---ACCTCGAAGTCTCTGTGAAA 732
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 SerThrAsnThrGluLeuGlnThrHisTyrValPheProLeuTyrValLeuLeuAlaArg 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 733 GTTCCCAACAAAAGAAAGATGTAAGTTGTCATATAGCGCAAGTTCCACA----- 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 ProThrSerAsnLeuSerLeuGluGlyHisSerProIleTyrArgPheSerArgValCys 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 787 -----GGTAAAGAGCAG-----GTG 801
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 LeuLeuThrSerPheSerGluHisGlyAsnLysAspAsnSerGluAlaThrPheIleIle 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 802 CTTTGTATTCCTGACCTCAATCAACAAAACCGGAAATTTCCGTCCTCTGCGAGTTTCC 861
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 157 ProAspValLysSerLeuSerThrSerArgAlaCysAsnHisAspIleIlePheIleSer 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 862 AGTAATGAATTTGAACCTAGTAACGCCATATGTTGAAGTCTTACTCGTCTATTAGA 921
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 CysGlyGlnVal----- 180
QY 922 GTGACTCGTCCAGGAAGAGAGAGTTAATGGAATGATTAATGAGAAACCAATGAAAT 981
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 -----GlyGlnSerAsnGlyGluAspAsnCysSer 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 982 ATTGATGTCATGAAGAGCTTCCAGCCAGAGAAACGAATCTGAGGATGGGGAAG 1041
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 -----GlyAsnHisValGluAspSerSer--- 198
QY 1042 ACATTGTTGCACAAATGACAGATATTTGATAAAACAGCGCTTTACAGCTTTTATAGTGGG 1101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 -----LeuGlnMetLeuGluGly 204
QY 1102 GAATATGAAGTAGCCATGCGAGAAATGGAAGATGTCCATGAAGAAAGAAAGCAACA 1161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 -----LysCysSer 207
QY 1162 TGG-----GAGACTATTCTTGATGGGAAGAGGTGCTCCATTCGAAACATTTTCT 1212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 TrpGlyLysIleProThrAsnLeuLeuAlaSerSerLeuGluSerCysValAsnLeuSer 227

```

```

QY 1213 CAGGACCTACGTTGCGATTCCTCTGTTGGACGAGGACCAATGATAAATCTACG 1272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 LeuGlyHisIleValGluLeuAlaSerLysValThrMetArgProSer----- 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1273 GTCCTTATTGCCAAACCTCTGCCACTAGAAATTCAGAGAGTCTC-----CAT 1320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 -----PheLeuGluProLysPheLeuGluGlnAspSerCysLeuThrPheCysSerHis 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1321 CAGGAAAACACAGCCTGTTTCACTTAACTACTCAAACTATTGCTGTTAAAGAACATTG 1380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 LysValAspAlaValGlySerTyrLysLeuGlnLeuCysMetSerAlaGlnIleAlaGly 281
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1381 ACTACAGATCTA-----CAAACAAGAAAAGAAAGATACCTCCAAATGAA 1425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 AlaArgAspMetSerLeuSerProTyrSerSerTyrAsnAspValProProSer 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1426 AACCGACAAAATTA-----AGAAATTTTATCAGTTTCTCTATTAAC 1467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 SerLeuSerAspIleIleArgLeuArgSerGlyAsnValLeuPheAsnTyrLysTyr 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1468 AACAAATACAGGCAACAAACTGAAGCAAGAGATGACCTGCATTGCCCTTGGTGTACTGTG 1527
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 322 AsnAsnThrMetGlnGluThrGluValThrGluAspPheSerCysProPheCysTyrVal 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1528 AACTGCGCAAACTTTATAGTTTCTCAAGCATCTTAACTCTGCCATAGAGATTATTC 1587
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 342 ArgCysGlySerPheLysGlyLeuGlyCysHisLeuAsnSerSerHisAspLeuPheHis 361
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1588 TTCAACTATGTTTATCATCCAAAAGTGCTAGGATAGATGTTTCTATCAAT---GAGTGT 1644
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 TyrGluPheTrpIleSerGluGluTyrGlnValValAsnValSerLeuLysAlaAspAla 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1645 TATGATGGTCTCTATCGAGAAATCTCAGGATATTTCATCGCCACCTGGATTCTCTTT 1704
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 382 TrpArgThrGluLeuPheAlaGluGlyValAspProArgHisGlnThr---PheSerTyr 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1705 -----AGTCCCAAGGACCCAGTTAAG---AGAACACCT 1734
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 ArgSerArgPheLysArgArgSerLysAsnThrMetGluLysIleArgHisVal 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1735 ATCACACATATTCTGTGTGTC---AGGCCAAAACCAACAAAAGCAGCATGTCTCAATTT 1791
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 HisSerHisIleMetGluSerGlySerProGluAspGluAlaGlySerGluAspAsnPhe 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1792 CTTGAATCTGAAGATGGG----- 1809
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 ValGlnGlyGluAsnGlyThrSerValAlaAsnAlaSerIleAspProAlaGlnSerLeu 460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1809 ----- 1809
Db 461 HisGlySerAsnLeuSerProProThrValLeuGlnPheGlyLysThrArgLysLeuSer 480
QY 1810 GAAGTAGAAGACAAAGAACATATAGTAGTGCCCAATCGTCTGTATTTCATAGTAT 1869
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 GluArgSerAspProArgAsnArgGlnLeuLeuGlnLysArgGlnPhePheHisSerHis 500
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1870 ACCTGCTTACCTCTCCGCCCAAGAAATG-----GAAGTAGATAGTAGAAGTCAAAAG 1923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 ArgAlaGlnProMetGlnLeuGlnValPheSerAspArgAspSerGluAspGluVal 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1924 GATCTCTGAATGGCTAAGAGAAAACCATTAACAAATTTGAAGATTTTCTGATTTAAT 1983
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 AspAspAspIleAlaAspPheGluAspArgArgMetLeuAspAspPheValAspValThr 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1984 GAAGGAGAGAGAGAGTGAATCTCGAATCTCCATGTCATGCAATGAGCATGGTATTAT 2043
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 LysAspGluGlyLeuIleMetHisMetTrpAsnSerPheValArgLysGlnArgValLeu 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2044 GTGCAATCAATGAATCATGCTGCTATGCTGTTTGTAGAAATTTATGGACAGAAAATA 2103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 AlaAspGlyHisIleProTrpAlaCysGluAlaPheSerGlnLeuHisGlyArgGlnLeu 580
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2104 ATTAGAAG---AATTATGTCGAAACTTCATCTCTAGTCAGCATGCATCACTTT 2160

```

[illegible]

RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phapthanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter C.J., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RN
RN SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaniniker J.S., Prochnik S.E., Smith C.D.,
RA Rupy J.L., Bergman C., Berman S., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommler B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RN
RN SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RN
RN SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RN
RN SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phapthanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003777; AAF57132.2; -
DR EMBL; AY052003; AAK93427.1; -
DR FlyBase; FBgn0039831; CG12054.
DR InterPro; IPR007087; znf_C2H2.
DR Pfam; PF00096; znf_C2H2; 3.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
KW Metal-binding; Zinc; Zinc-finger
SQ SEQUENCE 628 AA; 56610 MW; 860C67EA04D41CF5 CRC64;

Alignment Scores:
Pred. No.: 3,51e-09 Length: 628
Score: 227.00 Matches: 143
Percent Similarity: 34.57% Conservative: 81
Best Local Similarity: 22.07% Mismatches: 208
Query Match: 2.95% Indels: 216
DB: 5 Gaps: 26

US-09-874-162A-7 (1-4409) x Q3V9Z6 (1-628)
QY 70 GCTCTCTCTCTCCAAATACCTCCGATTCGGGGCTGGGACTCCACTTCCCCACCCCTG 129
Db 2 AlaValPheLeuIleAsnValCysLysTyrrAsnGlyCysGlyIleThrPheProSerLeu 21
QY 130 GCGGACCTCATCGAGCACATCGAGGACACACCATCGATACAGATCCACGGGTTTAGAA 189
Db 22 SerAspLeuIleSerHisIleGluAspThrHisIleAspTyrrAspProLysValValGlu 41
QY 190 AAACAGAGATTACAGAGCCACCTATGTTGCCCTGAGTACATACATAGATTATGACACA 249
Db 42 GlnLysGluGlnAlaGlnProLacCysLeuProLeuSerTyrrValLeuArgPheIleThr 61
QY 250 GATGTCGCCGCCGAGGAGGAGGAGTCCCTAAAGAGAGAGATTGAGCGGAGCTCTCGCTG 309

| | |
|------------------------|----------|
| Alignment Scores: | |
| Pred. No.: | 1,759-07 |
| Cred. No.: | 205-50 |
| Length: | 614 |
| Matches: | 135 |
| Percent Similarity: | 34.5% |
| Best Local Similarity: | 18.9% |
| Query Match: | 263 |
| Indels: | 203 |
| Gaps: | 31 |
| DB: | 10 |

US-09-874-162A-7 (1-4409) x Q7XSM1 (1-614)

| | | | |
|----|------|---|---------------------------------|
| QY | 409 | TCTTCTATCCGACGACGACCTCCGACAGAG----- | CCAACACAG 444 |
| Db | 11 | SerSerAspGlnGlnLeuAsnProGluGlnLeuAlaLeuTyrCysLysProLeuGlu 30 | : |
| QY | 448 | ATCTATAGATTTCTTCGAACTCGGAATCTCATACACCAATATTTTGTGCACAGAACTCTT 507 | : : : : : : : |
| Db | 31 | LeuTyrAsnPheIleArgHisArgAlaIleGluAsnProProTyrLeuGlnArgCysLeu 50 | : : : : : : : |
| QY | 508 | ACTTACATGTCTCATTCGAAATCTCCAGAACCAACATCAAAAGGAAAAACATTTAAAGTTGAT 567 | : : : : : : : |
| Db | 51 | LeuTyr----- | : : : : : : : |
| QY | 568 | GATATCTTATCAAAAGTAGAGAAATGAAGAGGACGACGAATCTCATAGCTTGTACGCT 627 | : : : : : : : |
| Db | 53 | -----LysIleArgAlaLysGlnLysLys----- | : : : : : : : |
| QY | 628 | CATTTGCAGCTTTACGTTTACTGGTTTCTTCACAAAAATGATAAGCCATCACCAAACTCA 687 | : : : : : : : |
| Db | 61 | ArgIleGlnIleThrIleSer----- | : : : : : : : |
| QY | 688 | GAANATGAA-----CAAAATCTGTTTACCTTGAAGTCTGCTTGTGAAAGTT 735 | : : : : : : : |
| Db | 72 | AsnAsnLysGluLeuGlnAlaGlnAsnIlePheProLeuTyrValLeuPheAlaArgPro 91 | : : : : : : : |
| QY | 736 | TGCCCAAAAAAAGAGAGATGTAAAGTTGGTCCAAATAGGCAAGTCTCCACACGTTAAAG 795 | : : : : : : : |
| Db | 92 | ThrSerAsnValProIleGluGlyHisSerProIleTyrArgPhe-----Ser 107 | : : : : : : : |
| QY | 796 | CAGGTGCTTGTGATTCCTGACCTCAATCAACAAAAACCCGGAAAT----- | : : : : : : : |
| Db | 108 | GlnAlaArgLeuLeuThrSerPheAsnAspSer-----GlyAsnAsnAspArgAlaGlu 125 | : : : : : : : |
| QY | 841 | -----TTCCCGTCCCTTGCAGTTTCCAGTAATGAATTCGAACCTAGTAAACAGC 888 | : : : : : : : |
| Db | 126 | AlaThrPheValIleProAspLeu----- | : : : : : : : |
| QY | 889 | CATATGTGTGAAGTCTTACTCGTCTTATTTAGAGTGACTCGTCCAGAAAGACAGAGTTT 948 | : : : : : : : |
| Db | 137 | IleAlaThrGlnAlaTyrGlyLeuThrPheIleLeuValSerArgGlyThrLysLys--- 155 | : : : : : : : |
| QY | 949 | AATGGAATGATTAATGAGAAACCAATGAAATATTTGATGTCATGAA-----GAGCTT 1002 | : : : : : : : |
| Db | 156 | -----AsnLysGlyArgThrGlyGlnAsnLeuCysGluAsnAspCysSerGluLys 172 | : : : : : : : |
| QY | 1003 | CCAGCCAGAAACAAATCGTGAGATGGGGAAAGACATTTGTTGCACAAATGACA 1062 | : : : : : : : |
| Db | 173 | HisValAspTyrSerSerLeuArgLysLeuAlaGlyLysCysPheTrpGlyLysIlePro 192 | : : : : : : : |
| QY | 1063 | GTATTTGATAAAACACGGCGCTTACAGCTTTTATAGTGGGAAATATCAAGTACGCCATCGAG 1122 | : : : : : : : |
| Db | 193 | IleThrLeuLeuAsnSerSerLeuGluThrCys-----AlaAspLeuIleLeuGly 209 | : : : : : : : |
| QY | 1123 | GAATTTGGAAGAAATGTCCAATAGCAAGAAAGACACATGGGAGACTATTTCTGTATGGG 1182 | : : : : : : : |
| Db | 210 | HisIleValGluSerProIleSer-----IleCysMetSerProGly 223 | : : : : : : : |
| QY | 1193 | AAGAGGCTGCCTCCATTCGAAACATTTTCTCAGGGACCTACGTTGCAGTTCACTCTTCGT 1242 | : : : : : : : |
| Db | 224 | TyrLeuGluProThrPhe-----LeuGluHisAspAsnCysLeuSerPheCysSerArg 241 | : : : : : : : |
| QY | 1243 | TGGACAGAGACCAATGATAAATCTACGGCTCTTATTTGCCAAACCTCTTGGCCACTAGA 1302 | : : : : : : : |

[illegible]

```

QY 2164 CTTATTAGCATATGCTCAATAGATAAAGCTGTTACCAAGCTCCGTGAATGCGACAAAA 2223
Db 598 LeuLeuSerAlaArgThrMetAspThrCysAsnArgIleLeuAspAspIle----- 604

QY 2224 TTAGAAAGGGGAATCGCTCCCTCGCAACGAA 2259
Db 605 -----LysAsnGluArgSerAspProLysLysGln 614

RESULT 4
Q81BL3 PRELIMINARY; PRT; 1605 AA.
AC Q81BL3
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN MAL7P1.127.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RA SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the ENBL/GenBank/DBJ databases.
DR EMBL; AL844506; CAD50987.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; P:protein kinase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot.Kinase.
DR Pfam; PF00069; pk_kinase_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 1605 AA; 194644 MW; B63DF776A3E206A2 CRC64;

Alignment Scores:
Pred. No.: 2.38e-06 Length: 1605
Score: 192.00 Matches: 263
Percent Similarity: 32.18% Conservative: 174
Best Local Similarity: 19.37% Mismatches: 429
Query Match: 2.50% Indels: 492
DB: 5 Gaps: 59

US-09-874-162A-7 (1-4409) x Q81BL3 (1-1605)
QY 508 ACTTACATGCTCATCGAACTCCAGAACACATCAAAAGGAAACATTTAAAGTTGAT 567
Db 233 ThrTyrLysGluAsnLysAsnValHis---TyrIleLysAsnLysLeuTyr-----Asp 249
QY 568 GATATGTTATCAAAAGTAGAGAAAAATGAAGAGAGACGAAGAATCTCATAGCTTCAGCT 627
Db 250 HisValIleSerLysAsnGluMetTyrThrGlyGlnThrThrLysAsnAsnTyrLeuAsp 269
QY 628 CATTGCGAGCTACCTTACTGCTTCTTCAC----- 660
Db 270 LysLeuPheLysAsnPheIleTyrPheTyrHisIleGlnThrLeuThrAsnValSerTyr 289
QY 661 -----AAAAATGATAGCCATCACCACAACTCAGAAAAAT--- 693
Db 290 AspGluArgThrLysGlnTyrLysGlnAsnLysLysCysIleLysAspLysLeuAsnAsn 309
QY 694 -----GAACAAAATCTGTTACCTCGGAAGTCTCTGCTGTG 729
Db 310 IleThrAsnAsnTyrLysGlnTyrAspLeuAsnSerPheSerAspSerAlaLeuTyrIle 329
QY 730 AAA-----GTTTGGCCACAAA---AAAAGAAAGGATGTAAGTTGTCAC----- 768
Db 330 LysGlnLysIleIleCysSerLysGluLysArgLysThrIleAsnIleLysGlySerHis 349
QY 769 -----ATAAGGCAAGTTCCACAGGTAAA 792
Db 350 LysTyrLysAspLysLysAsnGlnGlnIleTyrAsnGluLysAsnAsnGluThrArgGln 369

```

```

QY 793 AAGCAGGTGCTCTTTGATTCCTGACCTCAATCAACAAACCCGGAATTTCCCGTCCCTT 852
Db 370 LysGlnIleProLeuGluSerTyrLysLysLysThrLysLysIleAsnLysGluGlnGln 389
QY 853 GCAGTTTC-----AGT 864
Db 390 ThrIleGluGlnIleTyrGlnHisAlaLeuLeuValLysArgLysIleValAspIleGln 409
QY 865 AATGAATTGTAACCTAGTAACAGCCATATGTTAGAGTCTTACTGCTCTCTATTAGAGTG 924
Db 410 GlnGluGlnLysThrAspAsnLysTyrIleGluSerAsnTyrThr----- 424
QY 925 ACTCGTCCAGGAAGAAGAGAGATTAATGAATGATTAAATGCGAAGAAC----- 972
Db 425 -----AsnAsnValLysAsnAsnGlnThrAsnLysHisAsn 436
QY 973 -----AATGAAATATGATGTCATGAAGAGCTTCCAGCCAGAAAGA 1014
Db 437 AlaMetAspProSerAsnAsnGluAsnIleSerSerThrSerLeuAsnValSerGluLys 456
QY 1015 AAACGAAATCGTACGATGGGAAAGACATTTGTCACAAATGACAGTATTT----- 1068
Db 457 LysMetPheArgPhePhePhePheLysThrSerGluLysAsnAspThrLeuPheLysIle 476
QY 1068 ----- 1068
Db 477 GlnAsnValPheIleLeuIleIleTyrIleLysLeuPheTyrIleTyrTyrLeuLys 496
QY 1069 -----GATAAAACAGCGCTTACAGCTTTTAGTGGGATATAG----- 1108
Db 497 TyrTyrLysAsnLysLysAsnLysGluIleSerLeuLeuAsn-ValAsnIleGlnAspIle 516
QY 1109 -----AAGTAGCCATGTCAGG---AAATGGAAGATGTCCTCAATGAAGCAAGAAAGACACA 1161
Db 516 ePheLysArgLeuCysLysHisLysLeuAsnAsnTyrSerTyrProSerLysTyrArgHis 536
QY 1162 TGGGAGACTATTCTTGATGGGAAGAGCTGCTCCATTCG----- 1201
Db 536 shisIleIleAsnLysLeuLysArgLysArgTyrAsnAsnIleIleTyrGluIleIleLys 556
QY 1202 -----AAACATTTCTCAGG-----GACCTACGTTGTCAGTCTCACTCTCGT 1242
Db 556 sAsnIleIleGluTyrPheLeuAsnCysLysLeuAsnLeuSerPheIleLysLeuIleGln 576
QY 1243 TGGACAGGAGAGACCAATGATAAATCTACGGCTCTATTGGCCAAACCTTTGCCACTA-- 1300
Db 576 uAsnMetTyrSerGluPhePheAsnIleHisLeuLeuLysTyrAsnLeuArgAsnIlePhe 596
QY 1301 -----GAAATTCAGAGAGTCTCCATCAGGAAACAGCCGTGTTCAAGTAACTACTC-- 1354
Db 596 eTyrSerAspGlnAsnAspGlnAsnGlnLysThrLeuPheLeuAsnLeuLeuHis 616
QY 1355 -AAACTATTGCTGTTAAAGAATCATTTGACTACAGATCTACAAACAGAAAGAAAGGAT 1413
Db 616 sLysCysLeuLeuLeuAsnAsn----- 623
QY 1414 ACTCGAATGAAACCGCAAAATTAAGAATATTTTATCAGTTTCTCTATAACACAAAT 1473
Db 624 -LeuGlnCys--AsnSerThrSerLeuSerHisTyrTyrPhePheTyrAsnValAsn 642
QY 1474 ACAAGCAACAAACTGAAGCAGAGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1533
Db 643 IleTyrLysHisAsnAsnCysLeuGluThrAsnTyrThrGluLysAspThrPheAsnIle 662
QY 1534 CGCAAACTTTATAGTTTACTCAAGCATCTTAACCTGCGCATAGCAGATTTATCTCAAC 1593
Db 663 ProArgAsnMetSerLeuLeuGluAsnThrThrIleThrAsnGluLysValLeuMetAsn 682
QY 1594 TATGTTTATCATCCAAAGGTCAGGATAGATGTTCTTCTATCAATGAGTGTGTATGATGCG 1653
Db 683 LysValSerHisIleAsnAsnPheIleSerAsnAlaCysThrSerHisIleTyrThrThr 702

```

QY 1654 TCCTAT-----GCAGAAATCCTCAGATATTATTCATGCCAACT 1692
 Db 703 SerHisPheGlnAsnThrSerHisAsnAsnAsnAspAspAsnAsnAsn 722
 QY 1693 GGATTTGCTTTAGT-----CGCACGACAGTAAAGAACACCTATCACAT 1743
 Db 723 AsnLeuTyrPheAsnSerAsnLeuLysAsnGlyHisGlnAsnIleThrAspThrAsnAsn 742
 QY 1744 ATTCTTGTCGTGAGGCCAAACGAAAGCAAGCATGCTGCAATTTCTTGAATCTGAA 1803
 Db 743 AsnIleIle-----IleAsnThrGlu 749
 QY 1804 GATGGGAAGTAAAGACAAAGACATATAGTAGGCCCAATCGCTGTATTTCCAT 1863
 Db 750 SerGluArgValPheAsnAspIleLeuLeuAsnProAsnTyrAspAsnIleTyrPheAsn 769
 QY 1864 AGTGATACCTGCTTACCTCTCCGCTCCA-----CAAGAAATCGAAGTAGATGAA 1914
 Db 770 SerAlaAsnValArgGluMetProGluHisValGlnGluIleAsnLysLeuSerGln 789
 QY 1915 GATGAA----- 1920
 Db 790 GluGluLeuAlaLysLysAsnTyrAsnProTyrLeuValAspTyrIleLeuAsnVal 809
 QY 1921 -----AAGCATCTGAATGCTTAAGAGAAA- 1947
 Db 810 ThrLysTyrAsnGluTyrAspLysTyrPheIleAsnLysLysAspIlePheTyrTrp 829
 QY 1948 -----ACCATACAAATTTGAAGAGTTTCTCTGATGTTAAATGAA- 1986
 Db 830 PheHisMetLeuTyrAsnIleAsnTyrAsnGluLeuMetAsnValAsnSerPheLeu 849
 QY 1987 GGAGAGAAAGATGATGAA----- 2007
 Db 850 ArgSerProAsnIleLeuLysIleProTyrValLeuTyrLysArgLysAsnGlyThrTyr 869
 QY 2007 ----- 2007
 Db 870 PheAspMetSerLeuPheSerPheTyrLysLeuTyrLeuLeuLysLysTyrAspSerAsn 889
 QY 2008 ---CTCTGAATCTCCAT----- 2022
 Db 890 ThrLeuPheAsnValTyrGluAsnValSerThrPheLysLysLeuSerLysGluLysTyr 909
 QY 2023 -----GTCAATGAGCAT-----GGGTTTATGCTGCAATCAATG--- 2058
 Db 910 LysAsnIleLysLysIleLysHisGluValAsnTyrIleLysGlyAsnLysLysLys 929
 QY 2059 -----AATCATGCTGTATGCTGTTGTA 2082
 Db 930 SerValTyrThrCysValProArgLysArgCysGlyAsnTyrLysGluMetIleTyrIle 949
 QY 2083 GAAATTTATGACAG-----AAATAAATTAAGACAAATTTATGCGA 2124
 Db 950 SerThrTyrGluLeuLeuLysAsnArgPheThrValSerLysAsnMetLeuLys 969
 QY 2125 AACTTCATGCTTCATCTAGTCAGCATGATGATTTAATCTATTAGCATATATGTCATA 2184
 Db 970 LysTyrMet-----PheMetSerAsnAsnAspGlyAspIleCysAsnIleAsnLys 987
 QY 2185 GATAAGCTGTTACCAAGCTCGTGAATGCGCAAAATTAAGAAAGGGGGAATCTGCT 2244
 Db 988 LysLysGluIleThrGluLeuTyrGluAsnLysLysIleHisValTyrGluGlnGlu 1007
 QY 2245 TCCCTGCAACGAAAGAAATTAAGTGAAGCAAAATGGACAGCAATGATTTAGTCAA 2304
 Db 1008 TyrAsnIle-ThrLysLys----- 2304
 QY 2305 ATTAACCTCAAGAGAAAG-----CTTTGG-----AAACAGATAGTGTCTCA 2346
 Db 1014 sLysArgLysLysLysLysArgIleProLeuTyrGluAsnLysAsnValThrGlu 1034
 QY 2347 GGGGTTTCAACACAGAGCAAAACAAACTCT-----GAAAGCTCTAACCCCA 2397

Db 1034 nSer--AsnAsnLysHisLysSerMetAsnAsnCysThrHisGluLysThr--AspAs 1052
 QY 2398 TGTTATGACAAACACTGAAATTTACATTTTAGGAATTCATCTCTAAGAAATTTAT 2452
 Db 1052 nIleTyrAsnSerTyrAsnGluTyr-----AspAspLysAsnTyrAsnTyr 1067
 QY 2453 ---GTTTTTGT-----TTTAATCATATGTTCCAAACAGGCACTGT 2490
 Db 1067 rAspAspAspLeuPheIleSerSerAsnPhePheAsnIleSerPheAsnIleAsnAsn-- 1086
 QY 2491 TAGATGAAGTAATGATTTCAACAGATATTTGTATCAGGGTTCATCTTCATCTCATTA 2550
 Db 1087 -----LysHisSerTyrAlaLeuPheAsnAspTyrAsnLeuThrTh 1100
 QY 2551 TGCAGCATTACATGATATCATTGATGATCAATTAACAAATCTGTGACTTAAAGCA 2610
 Db 1100 rIleGlyIleTyrLeuLysGluPheTyr-GluIleIleLysAspAlaTyrSerPheIleA 1120
 QY 2611 TGAAGCAATATTTCAAGTATTTTAAACTCAACAAATGTCATCAAAATATGTTGAAT 2670
 Db 1120 rgarg-----AsnGluAsnPheIleAsn- 1127
 QY 2671 GATCTAGAAATTTATCATATATAATCAGAAATTTTTCATTTATGAACGGCTGTTTT 2730
 Db 1128 --ProLysSerSerCysIleTyrThrAsnSerPhePheIleTyrGln-----P 1144
 QY 2731 TCTACTTTGTAATTTGAGACATTTCTTGGGAGGAGAAATTTGGAATGTTCCCT--- 2786
 Db 1144 helysLeuHisIleLysTyrHisGluLeu-----LeuLysPheAsnProLeuA 1160
 QY 2787 -----TTTTAGAAATTTGAAGTGTCTTCATATGTCATCACTACAGAAA--- 2829
 Db 1160 spThrLeuThrLeuIleLysGluIleLysSerGlyTyrProCys--AsnMetArgAsnVal 1179
 QY 2830 -----CATTACTATTTGCAGTCAAACTTTGATCTCTGTTTT--- 2907
 Db 1220 LysAsnAlaIleIleTyrSerAsnAsnSerAsnValLysTyrCysMetTyrSerLys 1239
 QY 2908 -----TGAATCATTTTGTCAATTCGGAATGAAATTTATATATGT 2946
 Db 1240 LysTyrAsnHisLysHisAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 1259
 QY 2947 AATTTTACATTACAT----- 2961
 Db 1260 AsnArgLeuLeuTyrIleTyrLysAspLysTyrIleTyrAspAsnAspLeuIleGlySer 1279
 QY 2962 ---AAGTTCTCTTTTACAATTAATAAATAGCACTTCTTCATCTTATGCTGTTTGAAGA 3018
 Db 1280 LeuLysPheGlnLysLysLeuIleAsnLeuAsnGluTyrIleTyr--IlePheAsnSerL 1299
 QY 3019 TATTAATTTTCACTGTTGACAGTGAATG----- 3050
 Db 1299 euleuAsnTyrTyrIleProGluLeuSerMetTyrPhePheLysAsnAsnIleAsnIleI 1319
 QY 3051 -----CTATGTTGTTTATAAGATTAACAGACCA 3078
 Db 1319 leLysIleIleLysSerTyrIleTyrSerLeuPheCysThrPhePheAspLeuGlnAsnI 1339
 QY 3079 TTTGTTTTCATGATGATAATTTTATGTCATTTGCTACCCGGTATGTTTTTTTTTAAC 3138
 Db 1339 leTyrLeuLeuLeuAspAsnIleLeuIleGlnProSerSerTyrIle-IlePheIleCys 1358
 QY 3139 TTGAACATTTGCTGTTTGTGTTTCTTTTATTAATAGTAAATACACGGAAA----- 3192

```

Db 1359 IleSerIleLeuMetTyrLeuLysTyrLeuLeuAsnIleThrThrAsnGluAsnIle 1378
QY 3193 -----ATTAAGCTGTTCATATCTTTAAATAGGATTGCAACCAAGG 3234
Db 1379 TyrLysLysValPheSerLeuSerAsnLeuIleAsnLeuAsn-LeuIleLeuAsnThrSe 1398
QY 3235 AAAGAACCCATTGAGATTTAAAGATG-----CACTT 3267
Db 1398 rArgTyrLeuPheGlu-----ArgCysProLysTyrTyrLeuLeuPheProPheHisLe 1416
QY 3268 ATAAGGGAGAGAGTCTTCTTAAAGAGTCAACACAGAAACTGTTATGCTTTTATTGTTT 3327
Db 1416 uGlnIleGluGlySerSer-----IleAsnGluLysSerIle----- 1428
QY 3328 GCAAGGATGCTTTGTAAGTGTTCATGAATAGATATCAATCAATAGAGATAAGTGAATT 3387
Db 1429 -----LeuPheAspLysGlyGlyAsnGluAsnIleSer----- 1439
QY 3388 GAATCATTTTGAGCAATTTGCCCTGTTATGTTATGTTTCAACGCACATATTTCAGCTTG 3447
Db 1440 -AsnHis-----IleAsnTyrIleSerHis----- 1447
QY 3448 GATTTCCTCCACAGAAAGTGGATTCACTACTGG 3481
Db 1448 -LeuIleSerAsnLysLysTrpIleArgTyrTyr 1458

RESULT 5
Q813N9 PRELIMINARY; PRT; 9341 AA.
AC Q813N9;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF01120W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1] SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungel K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrrell B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Cordon C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagals K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929353; CAD51588.1; -
KW Hypothetical protein.
SQ SEQUENCE 9341 AA; 1121395 MW; BAE88CDEDF66B92D CRC64;

```

Alignment Scores:
 Pred. No.: 6.52e-06 Length: 9341
 Score: 188.00 Matches: 196
 Percent Similarity: 33.16% Conservative: 179
 Best Local Similarity: 17.33% Mismatches: 391
 Query Match: 2.44% Indels: 365

```

DB: 5 Gaps: 48
US-09-874-162A-7 (1-4409) x Q813N9 (1-9341)
QY 106 TCGGAGACTCCACTTCCCGCCGACCTCATCGAGCAGCATCGAGGACAAACCAATC 165
Db 2810 CysAsnArgHis-----LeuSerAsnTyrMetAsnSerIleLeuAsnAsnSerAla 2826
QY 166 GATACAGATCCACGGGTTTAAAGAAAACAAGAATTACAGACCAACCTATGTTGCCCTG 225
Db 2827 AspAsnLysHisAsnAsnLeuAspLysGluArgGluLysGluProIle----- 2842
QY 226 ACTTACATAAATAGATTATGACAGATGCTGCCCGCAGAGAGAGGAGTCCCTAAAGAG 285
Db 2843 ---PheMetAsnProTyrGluAsp-----LysAsp 2851
QY 286 AAAGATTCCGCGAAGCTCTCGCTGACTCTGTCTCAGTCTCAGTCTCTCGA-----GGGAAT 339
Db 2852 LysThrLysAspAspValGluHisSerIleAsnSerAsnGlnIleArgTyrValGlyAsn 2871
QY 340 GTGTCCACTCCCGCCGACGAGCAGTGTGAGAGCTTACTCCCGCTGACCCCAATC 399
Db 2872 ValLysCysIleGlyAsnAsnLysHisLysTyrValIleProPheLysSerAsnVal 2891
QY 400 ACCCGCTCTCTTCATTCGCGAGCAGCAGCTCCGACAGAGCAACACACAGATCTATAGATT 459
Db 2892 -----LysAsnLysMetGluGlyIleSerSerSerSerSerCysIleTyrLysLys 2909
QY 460 CTTGCACTCGGAATCTCATGACCAATATTTTTCAC----- 498
Db 2910 ThrLysArgLysLysTyrIleAspAspIleAspCysAsnAsnIleAsnAspLysLys 2929
QY 499 AGAATCTTACTTACATGCTCATCGAAATCCAG-----ACAAATCAAAAGG 549
Db 2930 LysSerTyrAspIleLeuTyrGlyAsnAsnSerGlnGluTyrTyrGluAsnIleLysAsn 2949
QY 550 -----AAACATTTTAAAGTTGATGATATGTTATCAAAAGTAGAGAAATGAAAGAGAGCAA 606
Db 2950 LeuArgAsnTyrLeuLysAsnAspMetIleLysAsnValLysLys-----SerSer 2966
QY 607 GAATCTCATAGCTTGTCTCAGCTCATTTGCAGTTTACGTTT----- 645
Db 2967 GluHisLysSerIleAspGlyAsnAlaAsnAsnLeuTyrAspIleProProAsnAspAsn 2986
QY 646 -----ACTGTTTCTTCCACAAAATGATAGCCATCACCACAACTCAGAAAATGAA 696
Db 2987 GluMetLysAspGluPheLeuIleLysAsnAsp-----AspAspAsnLys 3001
QY 697 CAAATTTCTGTTACCTCGGAGTCTGTTGGAAGTTTCCACAAAAGAAAGAGAT 756
Db 3002 MetAsnSerMetCysValAsp-----LysAspAspTyrLysGlnArgSerTyr 3017
QY 757 GTAAGTTGTCCAAATAAGG---CAAGTCTCCACAGGTAAAGAGCAG----- 798
Db 3018 IleSerTyrLeuLeuAsnTyrLysLysAsnGlyLysAsnGlyLysSerThrLysAsn 3037
QY 799 -----GTGCTTTGATTCTCTGACCTCAATCAACAAACAAA 831
Db 3038 AsnIleMetLysArgPheAsnSerLeuIleHisPheIleHisAsnAsnThrSer 3057
QY 832 CCGGA-----AATTTCGC 846
Db 3058 LysGlyAlaSerLysLysIleLysGlyLysArgSerLeuSerLeuPheSerPheLys 3077
QY 847 TCCCTTGAGTTTCCAGTAATGAATTAAGCTAGTAACAGCCATATCGTGAAGTCTTAC 906
Db 3078 SerLeuProIleArgArgLysPheGlyLysTrpAsnSerAsnThrIleAspValPhe 3097
QY 907 TCGTTG-----CTATTAGAGTGACTCGTCCAGGAAGAGAGATTAATCGA 954
Db 3098 GlnValGlyGlnGluAsnAsnTyrLysLysAsnLysLysLysThrLysLysPheGly 3117
QY 955 ATGATTAATGAGAAACC-----AATGAAATATTGATGTCAT 993

```


| | | | | | | | | | | | | | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 3337 | Asn | Phe | Met | Lys | Leu | Asn | Lys | Arg | Lys | Gly | Ser | Leu | Val | Cys | Cys | Tyr | Val | Pro | Val | Phe | 3350 |
| 2008 | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | 2022 |
| 3357 | Tyr | Lys | Asn | Ile | Ser | Leu | Thr | Tyr | Arg | Asn | Ile | Tyr | Lys | Cys | Leu | Leu | Ile | Leu | Leu | Asn | 3376 |
| 2023 | GT | CAT | GA | G | AT | GG | GT | TAT | TG | TG | C | AA | T | GA | T | GA | T | GA | T | GA | 2079 |
| 3377 | Met | Glu | Ser | Ile | Pro | Tyr | Leu | Ser | Lys | Leu | Tyr | Lys | Asn | Glu | His | Met | Cys | Lys | Tyr | Leu | 3396 |
| 2080 | GT | A | A | A | A | A | A | A | A | A | A | A | A | A | A | A | A | A | A | A | 2139 |
| 3397 | Phe | Gly | Asn | Lys | Asn | Lys | His | Asn | Leu | Asn | Asp | Asn | Tyr | Ile | Lys | Glu | Tyr | Ile | Leu | His | 3416 |
| 2140 | CT | AG | T | C | AG | C | AT | G | AT | C | T | T | A | T | T | A | T | A | T | A | 2199 |
| 3417 | Leu | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | 3423 | |
| 2200 | A | AG | CT | CG | C | T | GA | AT | TG | C | AG | CA | A | A | A | A | A | A | A | A | 2247 |
| 3424 | Glu | Lys | Glu | Gln | Tyr | Glu | Gln | Lys | Tyr | Ser | Phe | Glu | Leu | Lys | Lys | Asp | Ile | Phe | Lys | Asp | 3443 |
| 2248 | CT | G | C | A | A | C | G | A | A | A | T | A | CT | G | A | A | A | A | A | A | 2307 |
| 3444 | Met | Lys | Asn | Asp | Arg | Thr | Gln | Asn | Lys | Ile | Asn | Lys | Asn | Ile | ---- | ---- | ---- | ---- | ---- | 3459 | |
| 2308 | A | AC | T | C | A | A | A | C | G | A | A | A | G | C | T | T | T | G | A | A | 2355 |
| 3459 | e | Lys | Glu | Lys | Arg | Asp | Tyr | Tyr | Asn | Asn | Ser | Asn | Ile | Asn | Ile | Leu | Met | Lys | Tyr | His | 3479 |
| 2356 | A | A | C | A | G | A | G | A | A | A | A | A | A | A | A | A | A | A | A | A | 2415 |
| 3479 | s | Lys | Asn | Asp | Asn | Asn | Asn | Asn | Asn | Asn | Asn | Asn | Asn | Asn | Tyr | Asn | Asn | Asn | Asn | Asn | 3499 |
| 2416 | A | A | T | T | C | A | T | T | T | A | G | G | A | T | T | C | A | T | T | T | 2458 |
| 3499 | n | Lys | Asn | Asn | Tyr | Asn | Asn | Asn | Lys | Lys | Asn | Tyr | Ser | Tyr | Ser | Asn | Ile | Pro | Lys | Gln | 3519 |
| 2459 | - | G | T | T | T | T | A | A | T | C | A | T | G | T | T | C | A | A | A | A | 2517 |
| 3519 | u | Leu | Ser | Asn | Asp | Leu | Lys | Gln | Lys | Gly | Asn | Lys | Tyr | Tyr | Gln | Leu | Phe | Gln | Ser | G | 3539 |
| 2518 | A | T | A | T | T | C | T | A | C | G | G | T | T | C | T | A | ---- | ---- | ---- | ---- | 2537 |
| 3539 | l | n | Ile | Val | Gln | Asn | Ile | Phe | Asn | Tyr | Arg | Asn | Lys | Asn | Lys | His | Asn | Asn | Val | Asp | 3559 |
| 2538 | - | C | T | T | C | A | T | T | A | T | C | A | G | A | T | T | A | T | A | T | 2595 |
| 3559 | s | n | Ile | Glu | Asn | Val | Met | Gln | Asn | Glu | Val | Ile | Lys | Asp | Asp | Val | Lys | Pro | Leu | Asn | 3579 |
| 2596 | T | C | T | G | T | A | C | T | T | A | A | G | A | A | G | A | A | T | T | C | 2633 |
| 3579 | e | Cys | Ser | Leu | Tyr | Asp | Asn | Ser | Asn | Leu | Ser | Tyr | Pro | Phe | His | Ile | Tyr | Asn | Asn | Asn | 3599 |
| 2634 | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | 2679 | |
| 3599 | s | n | Asn | Asn | Asn | Asn | Gln | Lys | Cys | Asn | Ser | Ser | Asn | Ser | Asn | Asn | Cys | Asn | Ser | Leu | 3619 |
| 2680 | A | T | T | A | T | T | C | A | T | A | A | A | T | ---- | ---- | ---- | ---- | ---- | ---- | ---- | 2733 |
| 3619 | s | n | Tyr | Leu | Ile | Tyr | Lys | | | | | | | | | | | | | | |

```

QY 2908 TGAATCATTTTGTCAATTCGGAATGAAAAATATATAATTAATTTTACATTACATAAGTTC 2967
Db 3684 -----LeuSerThrIleAsnPheTyrPheLeuAsnAsnLeuAsnLysAsnLeuIle 3700
QY 2968 CTTTACATTAATAAATAGACACTTCTCACTTATGCTGCTGCTGAGAGATATAAT 3027
Db 3701 LeuCysAspIleLysGluThr-----LysAlaTyrPheLeu 3712
QY 3028 TTCACATTGTCACAGTGAATGCTAT 3054
Db 3713 HisLysTyrLysAsnIleArgCysTyr 3721

RESULT 6
Q8IEH5 PRELIMINARY; PRT; 3377 AA.
AC Q8IEH5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN MAL13PI.70.
OS Plasmidium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.,
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844509; CAD52282.1; -
DR InterPro; IPR008938; ARM.
KW Hypothetical protein.
SQ SEQUENCE 3377 AA; 404738 MW; E197442C1746178C CRC64;

Alignment Scores:
Pred. No.: 8,75e-06 Length: 3377
Score: 185.50 Matches: 226
Percent Similarity: 35.19% Conservative: 165
Best Local Similarity: 20.34% Mismatches: 381
Query Match: 2.41% Indels: 339
DB: 5 Gaps: 53

US-09-874-162A-7 (1-4409) x Q8IEH5 (1-3377)
QY 493 TTGCACAGACTCTTACTTACATGCTCTCATCGAACTCCAGAACAAACATCAAGGAAA 552
Db 357 IleAsnAsnIleAsnTyrAsnAsnHisAsnAsnAsnHisAspGluIleLysLysLys 376
QY 553 ACATTAAAGTTGATGATGTTA-----TCAAAAGTAGAG 588
Db 377 HisHisPheSerAspLeuLeuLysThrAlaThrGluGluLeuValSerLysIleSer 396
QY 589 AAATGAAGGAGAGCAAGATCTCATGCTTGTGTCAGCTCATTTGCGAGCTT---ACGTTT 645
Db 397 LysIleLysThrAspGluLysGlyGlyAspValGluGlyTyrMetLysValGluCysIle 416
QY 646 ACTGTTTCTTCCACAAAATGATAAGCCATCACCAACTCAGAAAATGAACAAAATTC 705
Db 417 AspGlyPheValPheGlnAsn-----TyrAsnGluTyrAsnLys 429
QY 706 GTTACCTCGGAAGTCTCTGTTGAAAGTTTGCACAAAAGAAA 750
Db 430 -----LysIleLeuArgLysMetLysAsnValLeuPheIle 441
QY 751 -----NAGATGTAAGTTGTCCCAATAGGCAAGTT 780
Db 442 TyrAsnPheAsnSerPheLeuSerTyrPheLysAsnIleLeuLeuAsnIlePheSer--- 460
QY 781 CCCACAGGTAAGAGCGAGTGCCTTTGATTCCTGACCTCAATCAACAAACAAACCCGGAAT 840
Db 461 -----AspValHisHisAsnTyrAsnGluGluLeuValValLeuHis 473

```

```

QY 841 TTCCCTGCTCCCTGTCAGTTCCAGTAATGAATTTGAACCTTAGTAACAGCCATATGGTG--- 897
Db 474 PheLeuAsnIleLeuGluLysPheAsnTyrTyrLysLysGluSerHisProPheLeuAsp 493
QY 898 -----AAGTCTTACTCGTCTGCTATTAGATGACT 927
Db 494 LysHisAspTyrSerSerTyrValAsnLysLysSerTyrAspLysLysPheGluPheLeu 513
QY 928 CGT---CCAGGAAGAAGAGAG--TTTAATGGAATGATTAATGGAGAACCAATGAAAT 981
Db 514 ArgAsnProAlaAsnThrAsnIleIleAsnLeuThrTyrAsnLysGluThrAsnAspTyr 533
QY 982 ATTGATGTCAATGAAGAGCTTCCAGCCAGAGAAAACGAAATCGTGAGGATGGGAAAAG 1041
Db 534 IleAsnIleIleLeuGluTyrGluAsnLysLeuMetArgGluLysGluPheTrpLeuTyr 553
QY 1042 ACATTTGTTGCCAAATGACACTATTGATAAAACAGG----- 1080
Db 554 SerGluLeuValProLeuThrValGluAspLysAsnAlaAsnLysPheTyrThrThrLeu 573
QY 1081 ---CGCTTACAGCTTTTATGATGGGAATATGAAGTAGCATCGCAAGAAATGGAAGAATGT 1137
Db 574 IleArgAsnThrLeuLeuLys-----ValIlePhePheGluSerPheLysLysCys 590
QY 1138 CCATTAAGCAAGAAAAGACCAATCGGAGACTATCTTGTATGGGAAGAGCGCTGCTCCA 1197
Db 591 IleLeuPheSerLysLysPheSerAlaLysSerIleIle-----LysIleIleGlu 607
QY 1198 TTGGAACACTTCTCTCAGGAGCTACGTTGTCAGTCTCACT----- 1236
Db 608 PhePheIleTyr-----IleLeuGluPheLeuCysSerThrCysValLysAsn 623
QY 1237 -----CTTGTGTCAGAGAGAGACCAATGATTAATCTACGGCTCCTATT 1281
Db 624 ArgAsnIleGlyGlyValPheTyrThrAspAsnIleSerAsnSerPheGluIleLysTyr 643
QY 1282 GCCAAACCTTTGCCACTAGAAATTCAGAGAGTCTCCATCAGGAAAACAAAGCCTGGTTCA 1341
Db 644 CysIlePheTyrPheThrLysIleIleLysGluLeuIleLysGlnArgLys----- 660
QY 1342 GTTAAACCTTACTCAAACTATTCTGTTAAAGAAATCATTTGACTACAGATCTACAAACAAG 1401
Db 661 ---LysAspThrTyrThrVal---ValLysGluLys-----GluAsnLysAlaArg 675
QY 1402 AAAGAA-----AAGGATCTCCAAATGAAACCCGACAA 1434
Db 676 PheGluGluValAsnLysMetTyrProProIleGluGluThrIleAsnGluAsnAsp 695
QY 1435 AAATTAAGAATATTTTATCAGTTTCTCTATACAAACAATACAAGGCAACAACCTCAAGCA 1494
Db 696 -----IleAsnAsnAspAspAsnIleAspSerAsnAsp 706
QY 1495 AGAGATGACCTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1554
Db 707 AspAspAsnIle----- 710
QY 1555 AAGCATCTTAAACTCTGCCATAGCAGATTATCTTCAACTATGTTTATCATCCAAAAGGT 1614
Db 711 -----Gly 711
QY 1615 GCTAGATAGATGTTTCTTATCAATCAGTGTATGATGCTCTCTATGACGAGAAATCCTCAG 1674
Db 712 SerAsnAspAspAspAsnIleGlySerAsnAspAspAsnAsnIleSerAsnAspAsp 731
QY 1675 GATATTTCAT-----CGCCAACCTCGA 1695
Db 732 AspAsnAsnIleSerAsnAspAspAsnIleGlySerAsnAspAspAsnIleGly 751
QY 1696 TTTGCTTTTAGTCGCAACCGCAGTTTAAGAGAACCTTATCACACATATTCTTGTGTGTC 1755
Db 752 SerAsnAspAspLysAsnAsnAsp-----SerAsnAspLysIleCys 765

```


RT falciparum.;
 RL Science 282:1126-1132(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=22255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow J.J., Suh B., Peterson J., Angiuoli S.,
 RA Perlea M., Allen J., Selengut J.J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrell B.,
 RA "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum.;
 RL Nature 419:498-511(2002).
 DR EMBL; AE001406; AAC71912.2; --
 DR PIR; C71610; C71610.
 KW Hypothetical protein.
 SQ SEQUENCE 2385 AA; 288299 MW; CC7CF642BC7EF0B9 CRC64;

Alignment Scores:
 Pred. No.: 1.09e-05 Length: 2385
 Score: 184.00 Matches: 236
 Percent Similarity: 33.20% Conservative: 183
 Best Local Similarity: 18.70% Mismatches: 362
 Query Match: 2.39% Indels: 481
 DB: 5 Gaps: 60

US-09-874-162A-7 (1-4409) x O96216 (1-2385)

QY 688 GAAATGAACAAATTCCTGTTACCTGGAAGTCTGCTGTTGAAAGTTTGCCACAAAAA 747
 DB 69 GLUASPcysAsnLysAsnMetThrIleAsnMetLeuLeuGlnAsnGluAsnLysLys 88
 QY 748 AGAAGAGTGAAGTGTCCAAATGAAGCAAGTCCACAGTAAAGACAGGTCCTTTG 807
 DB 89 IleileLysGluThrCysThrIleLysAsnValValThrAsnLysIleTyHisThrLeu 108
 QY 808 ATTCTGACCTCAATCAACAAACCCGGAATTTCCGTCCTCCCTGAGTTTCCAGTAAT 867
 DB 109 PheLeuValIleAsnLysHisThrHisAsnIleLeuCysSerLeuSerPheGluAsn 128
 QY 868 GAAATTGAACCTAGTAACAGCCATATGTTGAAGTCTTACTCGTTGCTATTAGAGTGACT 927
 DB 129 SerPheGluIleLeuAsnThrAsnPheValLysThr----- 140
 QY 928 CGTCAGGAAGAGAGATTAAATGAATGATTAATGAGAA-----ACCAATGAAAT 981
 DB 141 -----PheLysGlyLysIleLysSerMetAlaCysThrAsnAsnAsn 154
 QY 982 ATTGATGTCAATGAAGACTTCCAGCCAGAGAAAA-----CGAAATCGTGAG 1029
 DB 155 IlePheVal-----LeuIleLysLysLysLysLysLysLysLysLysLysLys 171
 QY 1030 GATGGGGAAGACATTTGTTGGCAAAATGACATATT-----GATAAAAC 1077
 DB 172 AsnGlnMetLysSerLysIleLeuAsnGlnAsnValLeuValSerLysHisThrLeuAsp 191
 QY 1078 AGCGCTTACAGCTTTAGATGGGGAATATAGTACCGATG-----CAGGAATGGAA 1131
 DB 192 ArgSerLeuLeuMetLysGlyLysLysLysLysLysLysLysLysLysLysLys 211
 QY 1132 GAATGTCCAATAGCAAGAAAGACCAATGGGAGACTATTCTTGATGGGAAGAGCGTG 1191
 DB 212 GluLysLysLysLysLysLysLys----- 219
 QY 1192 CCTCCATTGCAACATTTTCTCAGGACCTACAGTTGCGAGTTCACTCTTCGTGGACAGGA 1251
 DB 220 -----ThrAsp 221

QY 1252 GAGACCAATGATAAATCTACGGCTCCTATT---GCCAACCTCTTGGCCACTAGAAATTCA 1308
 DB 222 AsnLysAsnGluLysLysGlyHisMetGluIleLysAspValAsnGluLysLeuAsn 241
 QY 1309 GAGAGTCTCCATCAGAA-----AACAAAGCCTGGTTCAGTTTAAACCTACTCAAACTATT 1362
 DB 242 GluLysIleAsnGluLysAsnGluLysIleAsnGluLysAsnGluLysAsn 261
 QY 1363 GCTGTTAAAGATCATGCTACTACAGATCTACAAACAGAAAGAAAGAGATCTCAAACT 1422
 DB 262 GluGluLysAsnGluLysIleAsnGluLysAsnGluLysAsnGluLysLeuAsn 281
 QY 1423 GAAACCCGACAAAAATTAAAGATATTATTCATGATTCTCTATAACAACAATCAAGGCAA 1482
 DB 282 GluLysAsnGlu-----AspThrAsnLysAspPro 292
 QY 1483 CAAACTGAAGCAGAGATGACCTGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1542
 DB 293 TyrGluGluLysGluAsnAsp-AsnIleProLeuGlyAspHisHisSerValGlnTyrAs 312
 QY 1543 TATAGTTTACTCAAGCATCTTAACTCTCCATAGCAGATTTATCTCAACTATGTT-- 1600
 DB 312 nIlePheThrPheSerIleLeuAsnLysLysGluProAspLeuLysLysIleGlnPheSe 332
 QY 1601 ----ATCATCCAAAAGTGCTAGGATAGTGTTCCTATCAATGAGTGTATGATGGCTCC 1656
 DB 332 rAsnIleIleLeuProIle-----LysLysMetIleIleCysPr 345
 QY 1657 TATCAGGAAATCCTCAGGATATTATCGCCAACTGGATTGCTTTAGTCGCAACGGA 1716
 DB 345 cTyrAspGluLysIleIleIleLeuLeuSerHisLysSerIleValTyrIleThrAs 365
 QY 1717 CAGATTAAAGAACACCTATCACATATTCTGTGTGCGAGCCCAACAAACGAAACGA 1776
 DB 365 nLysAsnAsnAspLeuLysAsnMetPheIleIle-----LysGI 379
 QY 1777 AGCATGCTCGAAT----- 1789
 DB 379 uLeuIlePheAsnSerProIleIleThrThrThrTrpIleAspAsnTyrIlePheLeuI 399
 QY 1790 ----TCTTTGAAT-----CTGAAGATGGGGAAG----- 1813
 DB 399 eTyrPheLeuAsnAsnGluLeuIlePheLeuSerPheAlaLysProCysArgAsnLeuTy 419
 QY 1814 -----TAGAACAGCAAGAACATATAGTA----- 1837
 DB 419 rPheTyrLysCysIleAsnAsnTyr-SerHisIleThrSerPhePheTyrLysSerArgAs 439
 QY 1838 -----GTGCCACAATCGTCGTATTTCCATAGTGATA----- 1870
 DB 439 nLeuTyrIleSerPheLysThrLysGluIleValCysPheLysIleArgTyrTyrGluI 459
 QY 1870 ----- 1870
 DB 459 eProLeuThrValPheLysLysValAspLysAsnLysArgSerTyrIleGlnThrThrGI 479
 QY 1871 -----CCTGCTTACCTCTCCGTCCA 1890
 DB 479 uGlyAsnTyrIleAspAlaLysTyrLeuPheArgLysArgProArgTyrIleAsn-ThrA 499
 QY 1891 CAAGAAATGGAAGTAGATAGTAGTGAAGATGAAAGGAT----- 1926
 DB 499 snHisAsnGlnSer-AsnAlaLysAspLysAspGlyAsnAspValIleArgGluGlu 518
 QY 1927 CCTGAATGGCTTAAGAGAAAAACCATTTACAAATTGAAGAGTTTCTGTGATTGTAATGAA 1986
 DB 519 GluAspPheLeuArgAsnAsnAsn-----LysAsnPheSerAspValLysLys 534
 QY 1987 GGAGAGAAAGAGTGCATGAACTCTGGAATCTCCATGTCATGAAGATGGGTTTATTGCT 2046
 DB 535 A-GlyLysLysArg----- 538
 QY 2047 GACAATCAATGAATCATGCGCTGTATGCTGTTGTAGAAAAATTATGGACAGAAATAATT 2106


```
QY 1915 GATGAAGAGCTCTCAATGGCTAAGA---GAAAAACCATTTACACAAATGGAAGAGTTT 1971
Db 596 AspVallyHisCysTyrHisAsnMetGluLysPheIleTyrCysIleLysAsnBhe 615
QY 1972 -----TCTGATGTTTAATGAAGAGAGAGAAAGAGTATGATGAA 2007
Db 616 ProGlyProPheSerArgLysSerAsnLysValAspHisLysIleGluLysPheLeuLys 635
QY 2008 CTCTGGAATCTCATGTCATGAAGCAGTGGTTTATGTCGACAAATCAATGAATCATGCC 2067
Db 636 GlyTyr---IleAsnIleAsnLysSerArgSerValAsnAspAsn----- 649
QY 2068 TGTATGCTCTGTTGTAGAAATATGACACAGAAATTAATTAAGAAAGATTTATGTCGAAC 2127
Db 650 -----IlePheGluAsp-----IleGlnLysAsnCysLeuTyrAsn 661
QY 2128 TTCATGCTTCATCTAGTCAGC-----ATGCATGACTTTTAATCTTATTAGCATAATG 2178
Db 662 PheLeuLeuAsnIleLeuLysLysProTyrLeuThrAsnPheAsnLeuLysGluIle--- 680
QY 2179 TCAATAGATAAGCTGTTACCAAGCTCCGCGTAATGCGACAAA----- 2223
Db 681 SerIleGluLysLysAsnAspLeAsnLysAsnMetAsnHisLysGlyAsnAsnIleIle 700
QY 2224 -----TTAGAAAAGGGGAATCTGCTTCCCTCGCAAGAGAAATTAACCTGAAGAA 2274
Db 701 SerTyrPheLeuAspAsnTyrGlnGluLysLysAspIleLysAspAspIleTyr----- 718
QY 2275 CAAATGGGACAGCAAAATGGATTAGTGAATTAATCAAAAGAGAAAGCTTTGGAAACA 2334
Db 719 -----LeuSerAspAspCysAsnGluAspGluAspIleAspIle 731
QY 2335 GATAGTGTCTCAGGG-----GTTTCAAAACAG 2361
Db 732 AspSerLeuSerAsnAsnGluMetAspGluPheAspIleAsnGluTyrValLysAsnAsp 751
QY 2362 AGCAAAAACAAAACTCTGAAGAGCTCTAACCCCATGTTATGGACAAACATGAAAT-- 2419
Db 752 SerAsnLysMetLys-----LysTyrAspGluProCys-AsnAsnAsnHisIleAsnIle 769
QY 2420 -TACATTTTGGGAATTCATCTCTCAAGATATGTTTCTTTTATATCATATGTTCCA 2478
Db 769 eTyAsnProThrAsnAsnAsnSerAsnAsnIleMetAsnAsnAsn-ValAspLysL 789
QY 2479 AACAGGCACCTGTATAGT-----GAAGTAAATGATTTCAACA 2514
Db 789 yLysAsnValLeuAsnLysAspGlnLysLysCysTyrThrAspValTyrGluAsnSerL 809
QY 2515 AG-----GATATTGTATCAGGGTCTACTTCACTTCATTCATTCAGCAATTCATGT 2565
Db 809 yAsnIleAsnAspIleCysIleAsnAsnAsnThrTyrPhe----- 822
QY 2566 ATATCACCTTTTATGATGATCAATTAACATCTCTACTTTAAGCATGAAAGCAATATTT 2625
Db 823 -----PheAspPheIleAsn-----LysGluPheIleM 833
QY 2626 CAAAGATATTTTAACTCAACAAATGTCATCAAAATATGTTGAATGTATGATAGAAATATT 2685
Db 833 etGluLeuTyrLysLeuGluAsnLysAsnGluArgIleThrHis---AspAspIleTyrL 852
QY 2686 TCATATAT-----AAATCAGAAATTTTTCGATTTATG 2718
Db 852 euLeuTyrPheHisMetCysIleTyrAsnSerTyrLysAlaThrAsnValCysIleLysL 872
QY 2719 AACGGCTGTTTCTTACTTCTTAATTTGTGAGACATTTCTTGGGAGGGAATTTGGAAT 2778
Db 872 yGluLeuTyrLysTyrPhePheLeuValLeuArg-----Lys-TyrAsn 886
QY 2779 GGTTCCCTTTTGAAGATG-----AAGTGGTCTTCATAT-----GTCAAC 2820
Db 887 LysLysLysTyrTyrLysMetLeuAspLysTyrIleSerTyrIleLysArgSerIleAsn 906
```

```
QY 2821 TACAGAAAGCAAAAAATAGAAATTCAGAGATTTTATGAAATATATATGCAATTACT-- 2878
Db 907 AsnGluLeuGluMetLysAsnAspMetLysAsnIleTyr-LysIleTyrAsnTyrAsnIle 926
QY 2879 -----ATTGCACTCAAACTTTGATCCCTGTTTGTGAAATCATTTGTCATTCGG 2928
Db 926 eTyAspGluIleCysThrGluAlaPheValPhe-----PheLeuCysIleLeuAs 943
QY 2929 AATGAAAAATAT-----AATGTAATTTTACATCATCATAGTTCCCTTTTACAA 2976
Db 943 nLysLysAsnMetValPheLysLysAsnIleLeuValAsnTyrTrpTyrAlaPheTyrTh 963
QY 2977 TTAATAATAGCACTTCTTCATCTTATGCTGTTGAGAAGATATTAATTTTACATTC 3036
Db 963 rLeuIlePhe-HisAsnPheIlePheGlnTyrLysGluAsnGluIleGlnPheAlaAspT 983
QY 3037 TTGACAGTGAATGCTATGTTGTTTATTAAGATTA 3071
Db 983 yrlLysGluAspIleIleAsnPhePheIleTyrLeu 994

RESULT 9
O59904 ID O59904 PRELIMINARY; PRT; 1314 AA.
AC OS9904;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Silent information regulator 4.
GN SIR4.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98198828; PubMed=9539421;
RA Astrom S.U., Rine J.;
RT "Theme and variation among silencing proteins in Saccharomyces
RL cerevisiae and Kluyveromyces lactis.";
RL Genetics 148:1021-1029(1998).
DR EMBL; AF035007; AAC39438.1; -.
DR PIR; T09481; T09481.
SQ SEQUENCE 1314 AA; 146735 MW; 3C78CBB63319F8A6 CRC64;

Alignment Scores:
Pred. No.: 2,69e-05 Length: 1314
Score: 178.50 Matches: 188
Percent Similarity: 35.87% Conservative: 147
Best Local Similarity: 20.13% Mismatches: 360
Query Match: 2.32% Indels: 240
DB: 3 Gaps: 43

US-09-874-162A-7 (1-4409) x O59904 (1-1314)
QY 59 GCATCGCCGCCGCTCTTCTTCCTCAATPACCTGCCGATTCGGGGCTCGGACTCCACT 118
Db 110 SerSerArgGlnProSerProSerProSerProAlaAlaSerGlySerGlnTyrGlnLys 129
QY 119 TCCCCACCTGCCGCCACC---TCATCGACCATCGAGGACAAAC----- 159
Db 130 GlnLysSerTyrThrAsnArgLysSerSerThrSerHisThrGluMetLeuLysSerLeu 149
QY 160 -----CACATCGATACATACCA--- 177
Db 150 TyrValGlyAspSerLysValSerSerProGlnSerAsnIleAspLysAlaProGln 169
QY 178 -----CGGGTTTGTAGAAAAACAAAGATTA--- 201
Db 170 ArgLysProLeuGluArgThrGlyLeuSerLysArgValValGluAspGluValLeuArg 189
QY 202 ---CAGCAGCCCAACCTATGTTGCCCTCGAGTTACATAAATAGATTTCATGACAGATCTGCC 258
Db 190 ArgGlnGlnSerSerLeuSerSerSerGlyArgIleHisIleGluSerProPhe 209
```

| | | | | | | |
|----|------|--|------|------|--|------|
| QY | 259 | CGCCGAGCAGGAGTCCCTAAAGAAG-----AAGATTCCAGCGAAGCTCTCGCTG | 309 | 1210 | TCTCAGGACCTACGTTGCGAGTCTCTCTGTTGGACGAGAGACCAATGATAAATCT | 1269 |
| Db | 210 | GlnArgAsnSerLysProValSerLysSerIleValLysAlaProAsnLysValAsnPhe | 229 | 526 | ---GlnAlaAlaSerAlaSerGlyThrPheHisHisValAspGlu-----LysGlnSer | 542 |
| QY | 310 | ACTCTGTCCAGCTCAGTGTCT-----CGAGGGGAATGTCTCCACTCCCCCACCACAGC | 363 | 1270 | ACGGCTCTTATTCGCAACCTCTTGCCACTAGAAATTCAGAGAGTCTCCATCGAGAAAC | 1329 |
| Db | 230 | AlaPheGluThrSerLeuAsnAspIleAspLeuAsnLeuThrArgAlaAlaArgAlaSer | 249 | 543 | ThrGluGluThrAlaLysLeuLeuAlaLeuSerAspThrSer-----GluThrSer | 560 |
| QY | 364 | AGTGGAGCCTTACTCCCCCGTCACCCACCACCTACCCCTCTCTTCAATCCCGCAGC | 423 | 1330 | AAGCCTGTTCA-----GTTAAACCTACTCAAACTATTGCTGTGTTAAAGAA | 1374 |
| Db | 250 | SerGlyGly-----ThrProThrSerLysGlySerAla | 260 | 561 | SerAspGlySerSerSerGlyHisTyrValGlyAspPheGlnLeuLeuAsnThrMetLys | 580 |
| QY | 424 | AGCACTCCGACGAGCCACACAGATCTATAGATTTCTCGAACTCGGAAT----- | 474 | 1375 | TCATTCAGTACAGATCTACAAACAAGAAAGAGGATCTCCAAATGAAACCGCAAA | 1434 |
| Db | 261 | AlaValProAlaAlaPro-----LysThrLysAsnLysThrAsp | 273 | 581 | AlaArgMetValGlyProLysValAlaLysTrpLysAlaPhe-----AsnLysSer | 597 |
| QY | 475 | -----CTCATAGCACCATAATTTTTCGACAGAACTCTTACTTACATGTTCTCAT | 522 | 1435 | AAATTAAGAATATTTTATCAGTTTCTCTATTAACAACAATACAGGCAACAACCTGAAGCA | 1494 |
| Db | 274 | SerGlnValSerLysIleAlaAsnGluIleGlnAsnSerAsnLysSerArgIleSerLys | 293 | 598 | LysPheLysAspLeuMetLysTyrLeuIleAlaGluHisThr-----ProLeuAlaAla | 615 |
| QY | 523 | CGAACTCCAGAACAAACATCAAAAGAGAAACATTTAAAGTTGATGATATGTTATCAAAA | 582 | 1495 | AGAGATGACCTGCTGCTTGGCTTGTACTCTGAACTCCCGCAAACTTATAGTTA--- | 1551 |
| Db | 294 | AlaAsnSer-----ProIleGlnArgThrThrThrThrSerAsnIleProSerGlu | 311 | 616 | ArgSerAspValSerCysSerAsnAspIleCysAsnIleArgAsnTyrGluSerValGlu | 635 |
| QY | 583 | GTAGAGAAATGAAAGGAGAGCAAGAATCTCATAGCTTGTGAGCTCAATTTGCGAGTTACG | 642 | 1552 | CTCAAGCATCTTAAACTCTGCCATAGCAGATTTATCTTC----- | 1590 |
| Db | 312 | AlaLysGluLeu-----TyrAsnMetLeuSerArgLeuLysArgArg | 326 | 636 | ValHisHisLeuGlnLysLeuProArgLeuArgTyrLeuGluLeuLysProSer | 655 |
| QY | 643 | TTTACTGTTTCTCCCAAAATGATAGCCATCACCACAACTCAGAAAAATGAACAAAT | 702 | 1591 | ---AACTATGTTTATCATCCAAAAGTGCT-----AGGATAGAT---GTTTCTATCAAT | 1638 |
| Db | 327 | AsnValLeuPheAspLysGlyThrIleAsnSerGluAsnAsnAspLysThrGlyAla | 346 | 656 | AsnAsnPheIleSerLysProLeuArgAlaAsnAsnArgValGluLysLeuArgLysAsp | 675 |
| QY | 703 | TCTGTTACCTCGGAAGTCTCTGTGAAAGTTTGCACAAAAAAGAAAGGATGTAAGT | 762 | 1639 | GAGTGTATGATGCTCTCTATCGAGAAATCTCCAGGATATTCTCCGCAACCTGGATT | 1698 |
| Db | 347 | SerValLysValGluSerSerHisValGluIleIleGlySerLysArgAspHisThrHis | 366 | 676 | GluLeuTyrAsnGlySer-----LysGlnLysArgArgThrArgGlySer | 690 |
| QY | 763 | TGTCATAAGCGAAGTCCACAGGTAAAGCAGGTGCTTGTATCTCGACTCAAT | 822 | 1699 | GCTTTTAGTCGCAACGACCACTTAAG----- | 1725 |
| Db | 367 | ThrSerPheMetGluLeuProSerMetGlnAlaThrIleAlaAspLeuAlaLysThrThr | 386 | 691 | AlaProAspSerGluGlySerLysLysProLeuIleGluArgGlnIleGluAspGlyAsn | 710 |
| QY | 823 | CAAAACAAACCGGAAATTTCCCGTCCCTTCAGTTTCCAGTAATGAATTTGAACCTAGT | 882 | 1726 | -----AGAACCTATCACATATCTTGTGTGCGAGCCCAACGA | 1767 |
| Db | 387 | SerLeuValLysGluAsnSerAlaGlnIleGluValAspArgAsnLeuIleGlyProGlu | 406 | 711 | ValSerGluLysThrProGlnGluProThrThrSerIle-----ProGlyGlu | 726 |
| QY | 883 | AACAGCCATAGGTGAAGTCTTACTCGTGTGCTATTAGTAGTACTCGTCAGGAGAGA | 942 | 1768 | ACAAAAGCAAGCATCTCTGAATTTCTGAAGATGGGGAAGTAGAAGACGCAAGA | 1827 |
| Db | 407 | ArgSer-----ArgLysAsnGluLys | 413 | 727 | ThrAspValAsnGluLysValLeuGluProIleGluAsp-----IleGluMetLeuPro | 744 |
| QY | 943 | GAGTTTAATGAATGATTAAATGGAGAAACCAATGAAATATTTGATGTCAATGAAGAGCTT | 1002 | 1828 | ACATATAGTAGTGGCCACAATCGTCTGTATTTCCATAGTAGTACC-----TGC | 1875 |
| Db | 414 | GluPheGlnAspLeu-----GluThrArgThrSerSerProAlaLysGluSerLeu | 430 | 745 | ThrPheIleSerSerSerAsnArgLeuGluProAspAlaAsnThrLeuSerAlaHisAla | 764 |
| QY | 1003 | -----CCAGCCAGAGAAACGAAATCGTGAAGATGGGGAAG | 1041 | 1876 | TTACCT----- | 1881 |
| Db | 431 | AlaSerProSerAlaAspThrProGluArgLysValProGlnLeuAspLeuThrSerPro | 450 | 765 | LeuProSerSerValIleGlyValLysLysGlnProSerLeuThrThrValAsnLysPhe | 784 |
| QY | 1042 | ACATTTGTTGCACAAATCAGAGTATTGATATAAAACAGCGCTTACAGCTTTTAGATGGG | 1101 | 1882 | -----CTCCGTCCACAAAGAAATGGAAAGTAGATAGTAA----- | 1914 |
| Db | 451 | SerPheAspSerThrProGlyValThrGluLysMetAspLeuLysGluValProSerPro | 470 | 785 | ProAlaSerPheGluAsnGlyLeuGlnProAspAsnAlaGlnAlaSerThrGluIleGlu | 804 |
| QY | 1102 | GAATATGAA-----GTAGCCATGCGAGAAATGGAGAAATGTCCAAATAGC--- | 1146 | 1915 | ---GATGAAAGGATCCTGAATGGTGTAGAGAAAAACCATACACAAATCAAGAGTTT | 1971 |
| Db | 471 | GluTyrProSerSerAlaGlyGlnAlaValGlnGlnSerGluSerSerGlnIleSerHis | 490 | 805 | ThrAsnGluGlyAsnSerLysGlyLysAspGluLeuHisPheAspProGluLysCys | 824 |
| QY | 1147 | ----- | 1149 | 1972 | TCTGATGTTTAATCAAGAGAGAGAAAGATGATCAAACTCTGGAATCTCCATGTCATGAAG | 2031 |
| Db | 491 | ValIleGlnGlnAlaGluAlaAspSerGlySerHisHisGlyIleGlnValGluSer | 510 | 825 | GluAspValThrGlnPheArgLysGluTyrLeuAlaLeu---ArgAlaSerIleSerLys | 843 |
| QY | 1150 | AAAAGAGCAACATGGAGACTATTCTGATGGAGAGGCTGCTCCATTCGAAACATTT | 1209 | 2032 | CATGGGTTTATGCTGACAAATCAAAATGANTCATGCTGTGCTGTTGTTAGAAAATTAT | 2091 |
| Db | 511 | SerGluLeuAsnIleProThrIlePheGluLysLysAspAsnPro----- | 525 | 844 | AspAspThrCysSerSerSerLeuValAsn----- | 853 |
| | | | | 2092 | GGACAGAAAATAATTAAGAAAGAAATTTATGTCGAAAATCTTCATCTTCTAGTCAGCATG | 2151 |


```

Db      854  ---GluAspLeuAsnLysThrGluValIleGlnAsnLeuThrSerGluIlelle---Gln 871
QY      2152 CATGACTTTAATCTTATTAGCATATGTCATAGATAAAGCTGCTTACCAAGCTCCGTGAA 2211
Db      872 SerGluMetLysIleAlaSerLeuValGlyIleAsnHis-----GlnLeuArgGlu 888
QY      2212 ATGCAGCAAAATTAAGAAAG-----GGGAATCTGCTTCCCTGCTCAACAGAAATA 2265
Db      889 LysLeuGluGluLeuGluLysIleGlnAlaLysLeuValGlnGluLeuAspThrValLeu 908
QY      2266 ACTGAAGAACAAAT-----GGGACAGCAAAATGGATTAGTGAATAAATCACTCA 2313
Db      909 LeuGluSerLysGlySerPheSerAlaGlyThrAlaLysGlu---AlaGluValAsnAsn 927
QY      2314 AAAGAGAAAGCTTGGAAACAGATAGTGTCTCAGGGTTTCAACACAGAGCAAAACAA 2373
Db      928 LysAspLys-----ThrGluSerIleAsn-----AsnGluGlnAlaArgGlu--- 941
QY      2374 AAATCTGAAAAGCTCTAACCCCATGTTATGGACAAACACTG 2415
Db      942 -----LysSerLeuThrAlaAlaHisGlyLysLeulle 952

RESULT 10
Q8IBL9 ID Q8IBL9 PRELIMINARY; PRT; 1422 AA.
AC Q8IBL9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF07_0097.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844506; CAD50981.1; -
KW Hypothetical protein.
SQ SEQUENCE 1422 AA; 167923 MW; 0910F885329A8696 CRC64;

Alignment Scores:
Pred. No.: 3,276-05 Length: 1422
Score: 177.50 Matches: 174
Percent Similarity: 33.30% Conservative: 133
Best Local Similarity: 18.87% Mismatches: 219
Query Match: 2.31% Indels: 297
DB: 5 Gaps: 43

US-09-874-162A-7 (1-4409) x Q8IBL9 (1-1422)
QY 65 CCGCGCGCTCTCTCTCTCCA----- 85
Db 623 ProProProSerThrAspAsnArgPhePheAsnIleSerArgLysSerLeuLeu 642
QY 86 ATACTGCGGATTCGGGGCTCGGAGCTCCACTTCCCCACCCTGGCCGACCTCATCGAGC 145
Db 643 IleProAsnAspLysGlnMetLysIleSerAsn-----MetSerAsnThr 657
QY 146 ACATCGAGGAC-----AACCATCGAT-----ACAGATCCACCGGTTTGAAGAAACAA 195
Db 658 ThrGlyMetProTyrAsnHisValAsnMetMetAsnAsnProAsnMetIleAsnArgGln 677
QY 196 GAATTACAGCAGCAACCATGTTGCTGCTGATGTTACATAATATAGATTTCATGACAGATGCT 255
Db 678 Ser-----HisValSerLeuLeuAsnAsnThrPheThrGluPhe 690
QY 256 CCCGCCGAGAGCAG-----GATCCCTAAAGAAAGATTCAGCCGAGAGCTCTCGCTG 309
Db 691 GluMetLysAspLysAsnIleGluAsnSerAsnLeuAsnAlaProAsnIleAsnIle 710

```

```

QY 310 ACTCTGTCTCAGCTCA-----GTGCTCTCGAGGAATGTGTCC 345
Db 711 ThrSerAsnAsnAsnIleProProProAspThrAspIleAsnLysAsnLysMetIle 730
QY 346 ACTCCCCCACCACACAGCAGTGGAGCTTACTCCCCCGTGACCCACCACCATCACCCTCC 405
Db 731 AlaProProProTyr---AlaGlyAspAsnThrAlaSerValAsnProHisIleProAsn 749
QY 406 TCCTCTTCA-----TTCCGACGACGACTCCGACAGACCAACACACATCTAT 453
Db 750 PheHisSerGlyValMetAsnArgAsnSerIleSerProIleTyrProThrSerIle--- 768
QY 454 AGATTCTTCCGAATCTCATAGCACCACCAATATTTTTCGACAGAACTCTTACTTAC 513
Db 769 -----ArgSerAsnAsnMetIleAsnAsnMetMetAsnAlaAspGlyThrVal 785
QY 514 -----ATGCTCTCATCGAAATCCAGAACAAATCAATCAATCAATCAATCAATCAATCAAT 561
Db 786 LeuProAlaAsnMetAsnHisProAsnAsnMetLeuAsnVal----- 799
QY 562 GTTGATGATATGTTATCAAAAGTAGAGAAATGAAAGAGAGCAAGAAATCTCATAGCTTG 621
Db 800 ThrAsnAsnMetMetValSerProHisProMetAsnAsnGluMetLysAsnMetSerPhe 819
QY 622 TCAGCTCATTTGCAGCTTACGTTTACTGTTTCTTCCACAAAATGATAGCCATCCCA 681
Db 820 AsnLysAsnMetAsnIleSerLysThr-----Pro 829
QY 682 AACTCAGAAAATGAACAAAATTTCTGTACCTCGAAAGTCTGCTTGTGAAAGTTTGCCAC 741
Db 830 Asn-----AsnAsnValIleLeu----- 835
QY 742 AAAAAAAGAAAGATGTAAGTTGCTCAATY-----AGCAAGTTTCCCAAGGTAAA 792
Db 836 -----ThrLysAsnValAsnSerLeuMetAsnIleGlnArgGluProIleAsnSer 853
QY 793 AAGCAGGTGCTTTG---ATTCTGACCTCAATCAACAAACAAACCCGAAATTTCCCG--- 846
Db 854 TyrAsnAsnProLeuGlyPheLysProLeuAsnTyrAspAspProAsnSerIleProIle 873
QY 847 -----TCCCTTGCACTTTCCAGTAATGAATTT 873
Db 874 AsnTyrAspGluGlyAlaTyrPheAsnAsnAsnGlnMetAsnGlyThrAsnAsnAsnMet 893
QY 874 GAACCTAGTAACAGCCATATGTTGTAAGTCTTACTCGTCTGTTTATAGAGTACTCTCCA 933
Db 894 TyrHisMetAsnAspAsnIleLeuLysProAsnGluAspLeuIleArgIleSer----- 911
QY 934 GGAAGAGAGAGATTAAATGGAATGATTAAT-----GCACAATGACAGTATTGAT 1071
Db 912 GlyIleLysAspAsnAsnAsnAsnAsnAsnGluGluGluLysLysLysLys 931
QY 964 -----GGAGAAACCAATGAAATATTTGATGTCATGTAAGAGCTTCCAGCCAGAGAAA 1017
Db 932 GluLysGlyGluGluThrGluGlyGluGluValAspGluGln-----GlyGluIleGlu 949
QY 1018 CGAAATCTGTGAGGATGGGAAAAGACATTTGTT-----GCACAATGACAGTATTGAT 1071
Db 950 ArgAsnGlyGluGluAsnGluLysLysLysAspAsnLysAsnMetLeuIleGlnThr 969
QY 1072 AAAACAGCGCTTACAGCTTTTATAGTGGGAAATATGAAGTAGCCATGAGGAAATGAA 1131
Db 970 HisHisHisLysAspGluLeuLeuSerAsnArgAsnAsnPheMetAsnAsnGluLysLys 989
QY 1132 GAATGTCCAATAAGCAGAAAGAGCAACATGGGAGACTATTCTTGATGGGAGAGGCTG 1191
Db 990 GluLeuThrLysGluGluLysLeuHisGluTrp----- 1000
QY 1192 CCTCCATTCCGAAACATTTTCTCAGGACCTAGTTCAGTTCAGTTCCTTCTGTTGGACAGGA 1251
Db 1001 -----LeuGluTyrGluLeu-----ThrAsn 1007

```



```
QY 2348 -----GGTTTCAAAACAGACAGCAAAACCAAAACT 2378
Db 859 luHisIlePheAsnPheProIleGluTrpAsnIleProAsn---PheLysAsnAsnIleS 878
QY 2379 CTGAAAGCTTAACCCCATGTTATGGACAAACACTGAAATACATTTTAGGGAATTCAT 2438
Db 878 erThrLysLeuLysProIleTyLysLysIleThrGluTyIle-----GlyAlaIa 896
QY 2439 CCTCTAAGAATATGTTTCTTTTATCATATGTTTCCAAACAGGACACTGTTAGATGAA 2498
Db 896 spGluLysAspIleIleGluLysSerAsnTyPhe-ValLysGlnIleLeuAsnGlu 915
QY 2499 GTAATGATTTCAACAAGATATTTGTATCAGGGTTCTACTTCAC----- 2543
Db 916 ThrSerProLysAsnMetLeuValGluAlaGluLysPheLeuAspSerAspGlyLysIle 935
QY 2544 TTCAATATGACGACATTACATGATATACACTTTT 2576
Db 936 PheIleLeuAsnMetTyLysLeuIleIlePhe 946

RESULT 12
Q86H41
ID Q86H41 PRELIMINARY; PRT; 1693 AA.
AC Q86H41;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 416:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL, AC115598; AAO53202.1; -.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 1693 AA; 190577 MW; 25053531B44197C7 CRC64;
```

Alignment Scores:

| Pred. No.: | 6.35e-05 | Length: | 1693 |
|------------------------|----------|---------------|------|
| Score: | 174.00 | Matches: | 146 |
| Percent Similarity: | 22.97% | Conservative: | 129 |
| Best Local Similarity: | 17.51% | Mismatches: | 325 |
| Query Match: | 2.26% | Indels: | 234 |
| DB: | 5 | Gaps: | 31 |

US-09-874-162A-7 (1-4409) x Q86H41 (1-1693)

```
QY 187 GAAACACAGAAATACAGACCCCACTATGTCCTCGAGTTAC----- 231
Db 276 GlnGlnGlnGlnGlnGlnGlnGlnValSerIleThrLysGluTyGluLeuSerGluVal 295
QY 232 -----ATAAATAGATTTCATGACACAT-----GCT 255
Db 296 GluAspValLysIleGluGlnThrAspGlnGlnIleGluProIleLysSerSerIleGlu 315
QY 256 GCCCCCGAGAGAGGATCCCTAAAGAGAGATTCAGCCGAGCTCTCGCTGACTCTG 315
Db 316 AlaLeuGluThrLysGluThrAlaLysThrLysGluThrAlaGluThrThr 335
QY 316 TCCAGCTCAGTGTCGAGGGAATGTGTCCTACTCCCCACGCCACAGCAGTGGAGCGTT 375
```

```
Db 336 GluThrThrGluThrThrGluThrAlaGluThr-----LysGluThrAlaGlyThrThr 353
QY 376 ACTCCCCCGTGCACCCACCCATCACCCCTCTCTTCATTCCGCGAGCAGCTCCGACA 435
Db 354 GluThrThrThrThrThrThrIleThrGluThrAlaAlaIleValValGluGlnThrVal 373
QY 436 GAGCCACACAGATCTATAGATTTCTTCCAACTCGAATCTCATAGCACCATTATTTTG 495
Db 374 GluProAsnGlnGluIleThrValLysGlnGluLysGlu-----GluGluIleGlnGln 391
QY 496 CACAGAACTCTTACTTACATGCTCTCAT-----CGAAACTCCGACAAACAAC 540
Db 392 SerAsnGlnLeuSerLeuGluAspSerAspLysAsnGluIleLysSerAsnAsnSerAsn 411
QY 541 ATCAAAAGGAAACATTTAAAGTTGATGATGTTATCAAAAGTAGAGAAATGAAAGGA 600
Db 412 IleAsnArgAsp-----AspAspAspAspIleIle-----LeuProGln 424
QY 601 GAGCAAGATCTCATAGCTTGTGCTGCTCATTTGCGAGCTTACGTTTACTGGTTTCTCCAC 660
Db 425 GlnGluThrGlnSerSerSerSerLeuAspIleAsnLysThrAspLeuGlyPheAspAsn 444
QY 661 AAAAAATGATAAGCCATCACCAAACTCAGAAATGAAACAAATCTGTTACCTCGAAAGTC 720
Db 445 AsnAsnAsnAsnAsnAsnLysAsnSerAsnAsnLysAsnAsnAspIleIleGluSer 464
QY 721 CTGCTTGCGAAAGTTTGCACAAAAAGAGGATGTAAGTTGTCCTCAATAGGCAAGTT 780
Db 465 ThrIle----- 466
QY 781 CCCACAGGTAAACAGCAGTGCTTTGATCTCTGCTCAATCAACAAAAACCGGAAAT 840
Db 467 -----GlnThrSerSerIleProAspIleAsnSerAsnAspSerThrThr 481
QY 841 TTCCCGTCCCTTGCAGTT-----TCCAGTAATCAATTT 873
Db 482 ThrProThrLeuThrLeuThrSerProProThrThrAlaPheThrThrSerAsnGluLeu 501
QY 874 GAACCTAGTAAC-----AGCCATATGGTGAAGTCTTACTCTCGTT 912
Db 502 AspLeuSerAsnThrThrThrThrThrThrThrThrThrIleSerSerThrAspIle 521
QY 913 CTATTTAGAGTGACTGCTCCAGGAGAGAGAGATTAAAT-----GGAATGATT 960
Db 522 LeuAspAsnThrThrLysIleLysThrSerGluThrThrIleGluCysIleAspLysIle 541
QY 961 AATGGAGAAACCAATCAAAATATTCATGTCAT----- 993
Db 542 AsnAsnAspAsnThrAsnIleAsnThrAsnIleThrIleThrAsnIleAsnThrAsn 561
QY 994 -----GAAGAGCTTCCAGCAGAGAAAGAAACGAATCGT 1026
Db 562 ThrAsnIleThrIleThrAsnThrGlnThrGluLeuIleValGluSerAsnLysGluSer 581
QY 1027 GAG-----GATGGGGAAGACACATTTGTCACAAATGACA 1062
Db 582 GluProProProProLeuAsnAsnAspGluLysProThrThrThrProThrPro 601
QY 1063 GTATTTGATAAAACAGCGCTTACAGCTTTTAGATGGGGAATATGAAGTAGCCATGAC 1122
Db 602 IleAsnAsnGlnSerAsnGluIleAsnLysLeuGluAsnLys-----Glu 616
QY 1123 GAAATCGAAGATGCTCCATATAGCAGAGAAAGA-----GCAACATGGGAGACTATT 1173
Db 617 LysLeuGluAspAsnIleValSerLysGluLysIleIleValSerSerGluGluSerIle 636
QY 1174 CTGATGGGAGAGGCTGCTCCATTCGAAACATTTTCTCAGGCGACTACGTTGAGTTC 1233
Db 637 LeuThrAla----- 639
QY 1234 ACTTTCGTTGGACGAGGAGACCAATGATAAATCTACGGCTCCT-----ATTCCCAA 1287
Db 640 -----ThrAlaThrIleAsnAspLysIleThrAspGluAsnIleThrThr 655
```

```
QY 1288 CCTCTTGCCACTAGAAATTCAGAGAGCTCCATCAGGAAACAAACGCTGGT----- 1338
Db      |||      |||      |||      |||      |||      |||      |||
656 ProThrThrIleThrAlaSerAlaThrAlaGluGluLysGluAspSerGluThrIle 675
QY      |||      |||      |||      |||      |||      |||      |||
1339 -----TCAGTTAAACCTACTCAAACTATTGCTGTGTTAAAGAAATCATTGACTACAGT 1389
Db      |||      |||      |||      |||      |||      |||      |||
676 LeuGluGluSerIleLysProLysAspHisIleMetGluGluThrAspSerLysThrAsp 695
QY      |||      |||      |||      |||      |||      |||      |||
1390 -----CTACAAACAAAGAAAGAAAG 1410
Db      |||      |||      |||      |||      |||      |||      |||
696 AsnCysAsnAspIleSerMetGluAspAlaAsnSerAsnAsnAsnSerAsnAsnAsn 715
QY 1411 GATACCTCCAAATGAAACCGACAAATTAAGAAATATTATCACTTCTCTATACAAAC 1470
Db      |||      |||      |||      |||      |||      |||      |||
716 AspAsnAspAsnGluAsnAsnAsnAsn-----AsnAsnAsn 727
QY 1471 ATACAAGCCAAACAACTGAACCAAGAGATGACCTGCATTGGCTTGGTACTCTGAAC 1530
Db      |||      |||      |||      |||      |||      |||      |||
728 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn----- 740
QY 1531 TCCCGCAACTTATAGTTTACTCAAGCATCTTAACCTCTGCCATAGCAGATTATCTTC 1590
Db      |||      |||      |||      |||      |||      |||      |||
740 ----- 740
QY 1591 AACTATGTTTATCATCCAAAGGTGCTAGGATAGATGTTCTTATCAATGAGTGTATGAT 1650
Db      |||      |||      |||      |||      |||      |||      |||
741 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 760
QY 1651 GCTCCTATGAGGAAATCCTCAGGATATTCATCCCACTGGATTTGCTTTAGTCGC 1710
Db      |||      |||      |||      |||      |||      |||      |||
761 AsnGluLysSerGluAsnGluAsnGluInSerLysGlnAspThrLysSerGluSerAsn 780
QY 1711 AACGACCACTGTAAGAGAACCTATCACATATCTTGTTGTGCGAGCCCAACGAAACA 1770
Db      |||      |||      |||      |||      |||      |||      |||
781 AsnGlyThrAsnLysMetAsnGlyValAspHisSerPheIleAspAsnAsnGluLysAsp 800
QY 1771 AAAGCAAGCATCTCTGAA-----TTTCTTGAATCTGAAGATGGGGAATGAGAACAG 1821
Db      |||      |||      |||      |||      |||      |||      |||
801 AsnArgLysLeuThrGluGluGluValAlaLeuGluAsnThrGluGluAspIle----- 818
QY 1822 CAAGAACAATAGTAGTGGCCACAATCGTCTGTATTTCCATAGTAGTACTGCTTACCT 1881
Db      |||      |||      |||      |||      |||      |||      |||
819 -----AsnIleCysAspAsn 823
QY 1882 CTCCTGCCAAGAAATGGAAGTAGATGTAAGATGAAAGGATCCCTGAATGGCTAAGA 1941
Db      |||      |||      |||      |||      |||      |||      |||
824 GlnProProGlnGluGluGlnGluLysGluGluGluLysGluGluGluLysGlu 843
QY 1942 GAAAAACATTAACAATTAAGAGTTTCTGTGATGTTAATGAAGAGAGAAAGAGTG 2001
Db      |||      |||      |||      |||      |||      |||      |||
844 GluLys-----GluGlu-----GluLysGluGluLysLysLysGluVal 856
QY 2002 ATGAACCTCTGGAATCTCATGTCATGAAGCATGGTATTGCTGCAATCAATGAAT 2061
Db      |||      |||      |||      |||      |||      |||      |||
857 GlnLysGluGlyGluIleLeuAsnValLysGlu-----GluAsnThrGluGlu 872
QY 2062 CATGCTGTATGCTGTTGTAGAAATATTGGACAGAAAAATAATTAAGAAATTTATGT 2121
Db      |||      |||      |||      |||      |||      |||      |||
873 LysThrLeuGlnLeuGluIleLysLysIleGlnGluLysValSerLysLysMet--- 891
QY 2122 CGAAACTTCATGCTCCTAGCTAGCATGATGATCTTAATCTTATTAGCATATGTCA 2181
Db      |||      |||      |||      |||      |||      |||      |||
892 GluThrIleGlnGluMetValGlu---LysAspPheGluAspIle----- 906
QY 2182 ATAGATAAGCTGTTACCAAGCTCCGTGAATATGCAGCAAAATATAGAAAGGGGAATCT 2241
Db      |||      |||      |||      |||      |||      |||      |||
907 IleGluLysGluGluLysGlnIleLeuThrThrGlnGlnValSerLysGluAsnGlu 926
QY 2242 GCTTCCCTGCAACGAGAAATAACTGAAGA----- 2274
Db      |||      |||      |||      |||      |||      |||      |||
927 GluAsn---GluAsnGluLysValValGluSerIleIleIleLysLysGluIleGlu 945
```

```
QY 2275 CAAATGGGACAGCAAAATGGAATTTAGTGAATTAACATAAAAGAGAAAGCTTTGGAAACA 2334
Db      |||      |||      |||      |||      |||      |||      |||
946 GluAsnGluAsnGluAsnGluAsnLysAsnLysAsnGluThrThrPheGluLys 965
QY 2335 GATAGTGTCTCAGGGGTTTCAAAACAGACGACAAAACAAA 2376
Db      |||      |||      |||      |||      |||      |||      |||
966 ValAspIleAsnLysGluAsnGluGluSerLysGluAsnLys 979

RESULT 13
Q8IM18
ID Q8IM18 PRELIMINARY; PRT; 2033 AA.
AC Q8IM18;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF14_0073.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Bazzell B.,
RA "Genome sequence of the human malaria parasite Plasmodium
RA falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014817; AAN36685.1; -.
KW Hypothetical protein.
SQ SEQUENCE 2033 AA; 241130 MW; AA7F62DFF694870 CRC64;

Alignment Scores:
Pred. No.: 7,85e-05 Length: 2033
Score: 173.00 Matches: 165
Percent Similarity: 34.78% Conservative: 171
Best Local Similarity: 17.08% Mismatches: 333
Query Match: 2.25% Indels: 297
DB: 5 Gaps: 36

US-09-874-162A-7 (1-4409) x Q8IM18 (1-2033)
QY 154 GACAACCATCATCATACAGATCCACGGTTTGTAGAAAAACAAGATTACAGACGCAACC 213
Db      |||      |||      |||      |||      |||      |||      |||
654 GluAsnAsnIlePheAspLysSerMetIleThrGlnLysLysGluLysLysLysAsn 673
QY 214 TATGTCCTCGATGATACATAAATAGATTATCATGACAGATGCTCCCGCGAGAGAGGAG 273
Db      |||      |||      |||      |||      |||      |||      |||
674 PheHisGluTyrAsnIleGluAsnAsnHisLeuAsnAsnGluMethHisLysThrAspAsp 693
QY 274 TCCTTAAGAAAGAAATTACAGCGAAGCTCTCGCTCACTCTGCCAGCTCAGTGTCTCGA 333
Db      |||      |||      |||      |||      |||      |||      |||
694 AsnThrGluTyrLysLysLysLysGlnIleSerLeuLeuLysSerAsnIleLysHis 713
QY 334 GGAATGTGTCCACTCCCCACGCCACAGCAGGTGGAAGCTTACTCCCCCGTGACCCCA 393
Db      |||      |||      |||      |||      |||      |||      |||
714 AsnAsnLeuArgIleHisLysAsnThrAspLysGlyAsnAsnLysLysTrpIleSerGly 733
QY 394 CCATCACCCCTCTCTTCATTCGCCGACGACACTCCGACAGCAACACAGATCTAT 453
Db      |||      |||      |||      |||      |||      |||      |||
734 AsnIleIle---AsnAsnGluTyrLysThrThrLysAspIleTyrTyrAsnAsnLeuTyr 752
QY 454 AGATTCTTCGAACTCGGAATCTCATAGCACCACCAATATTTTGCACAGAACTCTTACTTAC 513
Db      |||      |||      |||      |||      |||      |||      |||
753 LysIleAsnArgValLys-----LeuGluLys 761
```

| | | | |
|----|------|---|------|
| QY | 514 | ATGTCCTCATCGAAATCCAGACAAAACATCAAAAGGAAAAACATTATAA----- | 561 |
| Db | 762 | GluAspAsnAsnAsnGlyArgIleArgPheLysArgGlyThrPheAsnGluThrTyrAsp | 781 |
| QY | 562 | -----CTTGATGATATGTATTCAAAAGTA | 585 |
| Db | 782 | AsnIleTyrHisHisLysLysCysTyAspLeuValLeuAspGluLeuLysValThr | 801 |
| QY | 586 | GAGAAATGAAGGAGACAGAATCTCATAGCTGTGCAGCTCATTTGCCAGCTTACGTTT | 645 |
| Db | 802 | ArgLysLysLysAsnLysLysHisTyrHisGluTyrAspLys----- | 815 |
| QY | 646 | ACTGGTTCTTCACAANAATGATAAG-----CCATCACCAAACTCAGAA | 690 |
| Db | 816 | -----MethHisAsnHisGluLysAsnIleLeuThrAsnAsnSerAsnAsnAsn | 832 |
| QY | 691 | AATGAACAAATATCTGTACCCTCGAAGTCTCTGTGAAAGTTTCCACAAAAAAGA | 750 |
| Db | 833 | AsnAsnHisAsnAsnAsnArgPheAspIleAsnLysMetLysGluProAspGluLysArg | 852 |
| QY | 751 | AAGGAT-----GTAAGTTGTCCAATAAGGCAAGTCCCACAGGTAAAAACGAG | 798 |
| Db | 853 | LysAsnAspLysSerSerIleThrCys----- | 861 |
| QY | 799 | GTGCCTTTGTATCTCGACCTCAATCAAACAAACCCGAAATTTCCCGTCCCTTGCAATT | 858 |
| Db | 861 | ----- | 861 |
| QY | 859 | TCCAGTAATGAATTTGAACCTAGTAACAGCCATATGTTGTAAGTCTTACTCGTGTATTT | 918 |
| Db | 861 | ----- | 861 |
| QY | 919 | AGAGTAGCTCGTCCAGGAAGAAGAGAGTTTTAATGGAATGATTAATGAGAGAACCAATGAA | 978 |
| Db | 862 | -----AspAsnGlu | 864 |
| QY | 979 | AATATTGATGTCATGAAGAGCTTCCAGCCGAGAAGAAA---CGAAATCGTAGGATGGG | 1035 |
| Db | 865 | AsnIleArgIleAspTyrLysArgSerAlaMetGluThrIleArgAsnGlnLys----- | 882 |
| QY | 1036 | GAAAAGACATTTGTTGACAAATCACAGTATTTGATAAAACAGCGCTTACAGCTTTTA | 1095 |
| Db | 883 | -----IleHisGlnValThr-----AsnGlnLeuAsn | 891 |
| QY | 1096 | GATGGGAATATGAATGAGCTTCAGAGAAATGGAAGAANTGCCAATAACGACAGAAGA | 1155 |
| Db | 892 | ValThrLysTyrGluSerLysLeuIleGluLeuGluLysMetIleLysLeuLysGlu | 911 |
| QY | 1156 | GCAACATGGGAGACTATTCTTGATGGGAAGGCTGCCTCATTCAGAAACATTTTCTCAG | 1215 |
| Db | 912 | LeuLeuLysLysGluAspIleGluAsnGluGlnLeuLysSerPheLeuIleLeuGlnAsp | 931 |
| QY | 1216 | GGACCT-----ACGTTTCAGTTCACCTTCGTTGGACAGAGACCAATGATATAATCT | 1269 |
| Db | 932 | GluGluIleAsnAlaLeuLysGluGluLeuLysIleLeuSerGluGlnAsnAsnLysLys | 951 |
| QY | 1270 | ACGCTCCTATTGGCAACCTCTTGGCCACTAGAAATTCAGAGAGTCCCATCAGGAAAC | 1329 |
| Db | 952 | -----LysLysIleSerLysThrLysLysAspLysSerHisGln-AsnThr | 966 |
| QY | 1330 | AAGCTCGTTTCAGTTAAACCTTACTCAACTA-----TTGCTGTTT | 1368 |
| Db | 966 | rThrPheTyrLys---AsnLeuIleLysSerLysProSerLysPheMetGluIleIleLe | 985 |
| QY | 1369 | AAAGAAATCATTGCTACAGATCTCAACAACAGAAAAAGGAAGGATACTCCAAATGAAAC | 1428 |
| Db | 985 | uLysAsn-----ValLysGluLysLysGlnGluLysLysThrCysAsnPhelysGluLe | 1003 |
| QY | 1429 | CGACAAAAATTAGAATATTTATFACGTTTCTCTATACAAACAATACAACGGCAACAACT | 1488 |
| Db | 1003 | uAspLysAsnAsnAspGluMetIle-MetLys----- | 1013 |

| | | |
|------|--|------|
| 1489 | GAAGCAAGAGATGACCTGCGCATTCGCCCTGGTGACTCTGAACTGCCGCAAACTTTATAGT | 1544 |
| 1014 | -----AspProIleIlePheAsp-----AsnLysAsnAsnIleAsnPro | 1026 |
| 1549 | TTACTCAAGCATCTTAACCTCTGCCATGACGAGATTATCTTCAACTATGTTTATCATCCA | 1608 |
| 1027 | IleValLysAsnValAsn----- | 1032 |
| 1609 | AAAGGTCTAGGATAGATGTTTCTATCAATGAGTGTATGATCGCTCTCTATGAGGAAAT | 1668 |
| 1033 | -----SerLeuSerThrSerCysAsnSerValHisSerCysTyrProTyrAsp | 1049 |
| 1669 | CCTCAGATATCATCGCCAACTGGATTGCTTTAGTCGCAACGACCGACCTAGTAAAGAGA | 1728 |
| 1050 | LysLysAsnIleGluAsnGluGluAsnGluLysAsnGluLysAsn----- | 1064 |
| 1729 | ACACCTATCACATATTTCTGTGTGTCAGGCCAAAACGACAAAGCAACGATGTTCTGAA | 1788 |
| 1065 | -----GluLysAsnGluLysAsnGluLysAsnGlu----- | 1074 |
| 1789 | TTTCTTGAATCTGAAGATGGGAACTAGAACACAAAGACATATAGTAGTGGCCACAAT | 1848 |
| 1075 | -----LysAsnGluGluAsnGluLysAsnGlnIleLysGln-----AlaAsnAsnHisAsn | 1090 |
| 1849 | CGTCTGTTATTCATAGTACTGCTGTTACCTTCCTCGTCCACAGAAATGGAAGTAGAT | 1908 |
| 1091 | GluThrAlaIlePheAsnAspMetProProAsnIleLysAsnAsnGlnIleAspMetAsn | 1110 |
| 1909 | AGTGAA-----GATGAAAAGATCCTGAATGGCTAAGGTAAGAGAAAACCC | 1950 |
| 1111 | AsnThrPheSerTyrIleCysCysAspMetArgLysArgSerLysAspSerSerThr | 1130 |
| 1951 | ATTACACAAATTCAGAGATTTCTGATGTTTAATGAAGAGAGAAA----- | 1995 |
| 1131 | GlyPheIleIleLysGluLysAsnAspGluAsnAlaSerHisLysLysLysThrMetPhe | 1150 |
| 1996 | -----CAAGTGATGAAAACCTCTGG | 2013 |
| 1151 | CysProMetTyrLysAsnGluTyrArgIleAsnLysAsnAspGluMetArgLysGluTyr | 1170 |
| 2014 | AATCTCCATCTCATGAGCATGGTTTATGCTGACAACTCAATGAATCATCGCTGTATG | 2073 |
| 1171 | HisProHisGluThrSerAsnGlyGlnIle-----HisAsnTyrValThrAsnLysTyrPro | 1189 |
| 2074 | CTGTTTGTAAGAAATATGACACAGAAAATAATTAAGAAGAAATTTATGTGCAAACTTCATG | 2133 |
| 1190 | IleLysAsnAspAsnThrLysIleAsnLeuLysValAsnIleSerGluGlyLysLys | 1209 |
| 2134 | CTTCATCTAGTCAGCATGATCTTAATCTTATTAGCATATATGTCATAGATAAAGCT | 2193 |
| 1210 | GluArgIleAlaSerVal-----AsnPheAsnLysCysGlu-----AspAspAsn | 1224 |
| 2194 | GTTACCAAGCTCCGTGAAATGACAGCAAAAATAGAAAAGGGCAATCTGCTTCCCTGCA | 2253 |
| 1225 | LysAsnAspIleGluAsnMetAsnGluArg-----ThrSerAsn-IleGln | 1239 |
| 2254 | AACGAAGAATAACTGAAGAACAAAATGGGACAGCAAAATGGATTATGTAATAACTCA | 2313 |
| 1239 | nThrMetAsnLysIleGluAsnGlnIleLysAsnLysIleAspCysAsnAsnMetGluAs | 1259 |
| 2314 | AAAGAGAAAGCTTTGGAACAGATA-----GTGTCCTCAGGGGTTTCAAAAC----- | 2359 |
| 1259 | nLeuHisLysTyrAsnLysGluSerAsnThrAsnAspAlaTyrHisAsnLeuAsnLeuCys | 1279 |
| 2360 | -----AGACAAAAAACAAAAAC-----TC | 2379 |
| 1279 | sProSerThrIlePheArgLysLysAsnAsnAsnLysLysAsnAsnValLysIleCysGln | 1299 |
| 2380 | TGAAGAGCTTAACCCCATGTTATGACAAACACTGAAATTCATATTGAGGAATCA-- | 2437 |
| 1299 | nGluLysIleAsnProHisIleVal-----HisIleAsnTyrIleSerAsnAspAsnPh | 1317 |
| 2438 | -----TCCTCTAAGAAATATGTTTGTGTT----- | 2461 |

```

Db      1317  eGlyGlnMetArgLysSerGlyLysAsnLysTyrPheLeuArgLysValLysGluGluLys 1337
QY      2462  -----
Db      1337  sLeaenGluGluLysLysGluAsnIlePheSerLysPhePheSerTyrIle-LysLysL 1357
QY      2482  AGGCACTGTAGATGAAGTAAATGATTTCAACAAGCATATT---TGATCAGGGTTCTAC 2538
Db      1357  ysSerileCysAsnGluLysLysAspHisAsnLysAsnLeuArgTyrAsnSerHisV 1377
QY      2539  TTCACCTTCATTATGCGAGCATTCATGATAT-----ATCACTT 2574
Db      1377  aHisLeuArgThrLysAsnTyrIleTyrLysSerProAspAsnLysArgSerGlnSerP 1397
QY      2575  TTATTCATGATCAATTAACCATCTGTACTTTAAGCATGAAAGCAATATTTCAAAGTATT 2634
Db      1397  heileasLeuPheLysAsn---CysThr-----AlaGlnAsnGlySerLysAsnT 1413
QY      2635  TTATAACTCAACAATGTCAATCAATATGTTGAATGATCTAGAATATTTTCATATATA 2694
Db      1413  yThrAspTyrValLeuThrAspAsnAsnPheAsnSerileAsnThrTyrGlnIleAsn 1433
QY      2695  AATCAGAA 2702
Db      1433  sPlysGlu 1435

RESULT 14
Q8IKP5
ID      Q8IKP5      PRELIMINARY;      PRT; 1163 AA.
AC      Q8IKP5;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Hypothetical protein.
GN      PF14_0559.
OS      Plasmodium falciparum (isolate 3D7).
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxID=36329;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=3D7;
RX      MEDLINE=22355705; PubMed=12368864;
RA      Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA      Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA      Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA      Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Anguoli S.,
RA      Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA      Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA      McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA      Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA      Fraser C.M., Barrall B.,
RT      "Genome sequence of the human malaria parasite Plasmodium
RT      falciparum."
RL      Nature 419:498-511 (2002).
DR      EMBL; AB014825; AAN37172.1; -.
DR      InterPro; IPR007087; Znf_C2H2.
DR      Pfam; PF00096; zf-C2H2; 1.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 1163 AA; 139588 MW; 6B37EB4621D0DAB1 CRC64;

Alignment Scores:
Pred. No.:      8.61e-05      Length:      1163
Score:          172.00      Matches:      195
Percent Similarity: 32.13%      Conservative: 135
Best Local Similarity: 18.99%      Mismatches: 336
Query Match:      2.24%      Indels:      361
DB:              5      Gaps:      54

US-09-874-162a-7 (1-4409) x Q8IKP5 (1-1163)
QY      106  TCGGAGTCCATCTCCCACTCCGACCTCGCCGACCTCATCGACCATCGAGGACCAACCATC 165

```

```

Db      271  CysGlyLeuArgPheMetThrSerMetAspLysLysLeuAsnHisLeuGlu---AsnHisTyr 289
QY      166  GATACACATCCACGGGTTTGTAGAAAAACAAGATTACAGCAGCCAAC--- 213
Db      290  LysLysSerGlnPheTyrLeuAsnAsnSerGluArgSerPheThrLysSerLysLys 309
QY      214  ---TATGTTCCTCCAGCTTACATAAAAT--- 237
Db      310  ArgPheLeuTyrLeuAspHisIleAsnLeuProIleGluValPheValCysLysAsnTyr 329
QY      238  -----AGATTTC 243
Db      330  SerilePheGluAspPheTyrAspAsnValValThrLysAsnMetAlaSerPheAsnPhe 349
QY      244  ATG-----ACAGATGTCGCCCGC 261
Db      350  GlnGlySerHisIleGluGluGlnIleLeuAspAsnAsnAspAsnSerTyrHis 369
QY      262  CGAGAGCAGGAGTCCCTTA-----AAGAAGAAGATTTCAGCCGGAAG 300
Db      370  AsnAspAspGluLysIleTyrTyrGluGluGlyLysLysLysLysAspGluSerLys 389
QY      301  CTCTCGCTGATCTGTCCAGCTCAGTGTCTCGAGGGAATGTCCCACTCCCGCCAGCCAC 360
Db      390  AsnAsnAspSerLeuTyrSerTyrLeu---HisThrLysIleHisThrHisProAsnLys 408
QY      361  AGCAGTGGAAAGCCTTACTCCCGCGTGACCCCATCACCCCTCCTCTTTCATTCGCGC 420
Db      409  AspLysAsn-----GlnGluSerPheLeu 416
QY      421  AGCAGCACTCGGACAGAGCCACACAGATCTATAGATTTCTCGAATCGGAATCTCAT 480
Db      417  AsnAsn-----AsnThrLysAsnIleGln 424
QY      481  GCACCAATATTTTGCACAGAACTCTTACTTACATGTCTCATCGAAACTCCAGAACAAAC 540
Db      425  GluLysTyrAsnValTyrAsnMetLeuSerTyrAsnAsnAsnAsnAsnAsnAsp 444
QY      541  ATCAAA-----AGGAAACATTTAAAGTTGATGATGATATGTTATCAAAAGTA 585
Db      445  ValAlaAsnLeuPheAsnPheLysLysThrTyrAsnAsnIleAsp----- 459
QY      586  GAGAAATGAAAGAGAGAGCAAGATCTCATAGTGTGTGCTCATTTGCGACGTTTACGTTT 645
Db      459  ----- 459
QY      646  ACTGGTTTCTTCCACAAAATGATAGCCATCAACCAACTCAGAAAATGACAAAATTCCT 705
Db      460  ---AsnPhePheHis-----AsnLysGluThrIleHisAsnMet 471
QY      706  GTTACCTCGGAAGTCTGCTGTGGAAGTTTGCACAAAAAAGAAAGATGTAAGTTGT 765
Db      472  IleAsnProAsnMetAlaValAspLys-----AsnLysLysLysArgAsnIle--- 487
QY      766  CCAATAAGGCAAGTTCCACAGGTAAAAGACAGGTGCTTTGATCTCTGACCTCAATCAA 825
Db      488  -----LeuAsnAsp 490
QY      826  ACAAAACCGGAATTTCCCGTCCTTCAGTTCAGTTCAGTAATGAATTTGAACTAGTAAC 885
Db      491  ArgGluLysGly-----SerAsnGluIleAsnPhePheAspTyrIleTyrGlyAsn 507
QY      886  AGCATATGTGAAGTCTTAC---TCGTTGCTATTTAGAGTGACTCGTCGAGGAAGAAGA 942
Db      508  GluAsnThrTyrAspValTyrLeuSerAsnTyrTyrThrValThrGluAspAsnSerVal 527
QY      943  GAGTTTAATGGAATGATTAATGAGAAAC---AATGAAAATATT----- 984
Db      528  LeuIleAsn---TyrIleAsnGlyProThrMetTyrArgAsnIleMetLysCysLeuGlu 546
QY      985  ---CATGTC---AATGAGAGCTTCAGCC-----AGAAGAAAACGAAATCGT 1026

```

Db 547 IleLysAspIleTyrAsnTyrLysPheProSerTrpIleProLysArgSerIleAsn--- 565
 QY 1027 GAGGATGGGAAAGACATTTGTCACAAATGACAGATATTGTAATAAACAAGCGGCTTA 1086
 Db 566 -----AsnPhePheIleLysArgIleIleGluIleAspLysAsp----- 578
 QY 1087 CAGCTTTTAGATGGGAATATGAGTAGCCATGCGAGAAATGGGAAGATGTCGAATAGC 1146
 Db 579 -----LeuIleGly-----IleGln 583
 QY 1147 AAGAAAGACGACAACTGGAGACTATTCTTGATGGGAAGAGGCTGCCCTCCATTCGAACA 1206
 Db 584 LysLysThrAsnIlePheThrAsnIleLeuSerGlyArg-----ThrProSerAsnThr 601
 QY 1207 TTTTCT---CAGGACCTACGTTGCGATTCAGTTCACTCTCGTGGCAGGAGAGACCAATGAT 1263
 Db 602 PheProLeuAsnSerTyrAsnIleGlnAspAspLeuAsnIleAsnAsnThrAsnAsn 621
 QY 1264 AAATCTACGGCTCTATTGCCAAACCTCTGCCACTAGARATTGAGAGGCTC----- 1317
 Db 622 AsnLeuLysGlyLysAsnLysLysLysIleAspValIleAsnLysAsnAsnLeuTrpAsn 641
 QY 1318 -----CATCAG---GAAACCAAGCGCTGTTCCAGTTAAACCTTACTCAAACTATT 1362
 Db 642 IleCysThrThrHisGluPheTyrAsnSerLysGlyLeuCysProIleAsnValPheLeu 661
 QY 1363 GCTGTTAAGATCATTCAGTACATCTACATCTACAACA-----AGA 1401
 Db 662 SerTyrGlnLysGluAsnThrLeuAspHisGlnThrGluHisSerAspLysAsnAspHis 681
 QY 1402 AAAGAAAGGATACCCAAAT-----GAAAC 1428
 Db 682 AspAspLysGluSerAsnAsnTyrProMetGlnAsnSerIleCysThrGlnGluSer 701
 QY 1429 CGACAAATTAAGAAATTTTATCAGTTCTCTATACAAATACAAAGCAACAACT 1488
 Db 702 ValGluAsnAsnHisValLeuTyrSerPheLeu-----AsnIlePheAsnGlnLysTyr 719
 QY 1489 GAAGCAAGAGATGACCTGCATTCGCTGGTGTACTGTAAGTCCGCAACTTTAT--- 1545
 Db 720 PheAlaGlnThrAspIlePheLysAsnIleLeuPheHisIleArgLysHisTyrPhe 739
 QY 1546 AGTTTACTCAAGCAFTCTTAA-----CTCTGCCATAGCAGATTTATC 1587
 Db 740 MetIleIleAsnGluLeuLysAspAsnIleAspPheAsnGluLeuHisSerArg---Leu 758
 QY 1588 TTCAACTATGTTTATCATCCAAAGGCTGCTAGGTAGATGTT----- 1629
 Db 759 PheGlnTyrIleTyrAsnAsnTyrSerLysThrIleSerIlePheAsnIleAsnTyr 778
 QY 1630 TCTATCAATGATGTTATGATGGCTCCTATGAGAAATCCTCAGGATATTTCATCGCAA 1689
 Db 779 LysIleAspThrCysPhe-----LeuCysLys 787
 QY 1690 COTGATTTGCTTTAGTCGCAAGGACCATTAAGAGAACACCTATCACATATCTT 1749
 Db 788 GluAsnPheSerPheGluTyrSerTyrGluTyrAsnAspPheTyrTyrThrAsnValile 807
 QY 1750 GTGTCAGGCCCAACGAAACGAAAGCAAGCATGCTGAATTTCTGATCTGAAGATGGG 1809
 Db 808 CysValAspLeuLys-----AsnValTyrGluAsnAspAspValGluAspThr 823
 QY 1810 GAAGTAGAACGACGAAGACATATAGTAGTGGC----- 1842
 Db 824 GluGluAspAlaAsnLysAsnTyrAspIleAspMetSerIleLysArgIleAspHisMet 843
 QY 1843 -----CACAACTGCTGTATTTCCATAGTATCTGCTTACCTCCTC 1884
 Db 844 CysAspGluTyrValTyrAsnAsnAsnSerTyrAspIleMetAspLysCysLeuTyrGlu 863
 QY 1885 CGTCCACAGAAGATG---GAAGTAGATAGTGAAGTGAAGGATCTGGAAGTGAAGA 1941
 Db 864 LysHisLysGluLeuAsnGluLeuLeuTyrGluAsnAsnLysGlu-----MetSer 880

QY 1942 GAAAAAACCATTACACAAATTCGAAGAGTTT-----TCTGATGTTTAA 1983
 Db 881 AspAspAlaIleThrGlnValGluLeuLeuAsnLysMetValAsnAsnGluIleAsn 900
 QY 1984 GAA-----GGAGAGAAAGACTGATGAACCTCTGGAATCTCCATGTCATGAAGCAT--- 2034
 Db 901 GluCysSerAsnHisAsnAsnIleLysLysValAsnAsnIleAsnThrTyrAspAsnIle 920
 QY 2035 -----GGTTTATTGCTGACAAATCAATCAATGAATCAT 2064
 Db 921 AsnIleAsnAsnIlePheThrAspIleLysLysIleAspAspAsn----- 936
 QY 2065 GCCTGTATGCTGTTGTGAGAAAATTTATGACAGAAAATTAATTAAGAAATTTATGTGCA 2124
 Db 937 ---CysPheLeu-----GlyAsnThrIleIleGluSerAsnMetAspCys 950
 QY 2125 AACTTCATGCTTCATCTAGTCAGCATG-----CAT 2154
 Db 951 AsnPheHisMetAsnIleIleAsnLysLysAsnLysLysAspHisIleMetAsnHis 970
 QY 2155 GACTTTTAATCTTATTAGCATAATGTCATAAGTGTTCACCAAGCTCGTGAATG 2214
 Db 971 TyrAsnAsnLeuPheLysAsnLeuCysIleProTyrAspThrLeu-----AspIle 987
 QY 2215 CAGCAAAATTAAGAAAGGGGAATCTGCTTCCCTGCAACGAAAGAAATTAACAGAA 2274
 Db 988 LeuHisArgIleLysLysGly-----AspIleThrGlyLys 999
 QY 2275 CAAATGGGACGACGAATGATTTAGTGAATTAATCAAAAGAGAAAGCTTTGAAACA 2334
 Db 1000 ThrAsnMetThrThrAsnAsnPheSerAsnAsnThrThrAsnAspGluGlyTyrThrAsn 1019
 QY 2335 GATAGTGTCTCAGGGGTTT----- 2353
 Db 1020 GluGluIle-AspAsnPheLeuLysAsnTyrAspAsnIleThrAsnAspIleLysAsnAl 1039
 QY 2354 -----CAAAACAGACGACAAACCAAAACCTCTGAAAAGCTTAACC 2394
 Db 1039 aValPheGlnSerThrTyrHisAsnArgSerProSerLysAsnLysAsnSerPheAsnAs 1059
 QY 2395 CCATGTTATGCAACACACTGAAATTAATTTAGGAAATTCATCTCTAAG----- 2446
 Db 1059 PileIleMet-----SerTyrIleTyrAsnAspGluAsnLysLysLysGlnHi 1075
 QY 2447 ---AATTATGTTTGTGTTTAAATCATATGTTCCAAACAGGCACTGTTAGATGAAGTAA 2502
 Db 1075 sSerAsnIlePhePhePro----- 1081
 QY 2503 ATGATTTCAACAGGATATTGTCATCAGGTTCTCTACCTTCATTTATGACGATTAACA 2562
 Db 1082 -----SerLysAsnTyrThrTyrAsnPheThrTyrPheHisIleGlnCysPheLys 1099
 QY 2563 TGTATATCATTTTATGATGTCATTAACAACTCTG-----TACTTT-----AAGCA 2610
 Db 1099 sAsnTyrIleGluTyr-----AsnIleLeuProTyrTyrPheLeuThrLysLeu 1115
 QY 2611 TGAAGAGCAATATTTCAA 2629
 Db 1115 uAsnAspLysTyrPheLys 1121

RESULT 15

Q810T3 PRELIMINARY; PRT; 1396 AA.
 ID Q810T3;
 AC Q810T3;
 DT 01-JUN-2003 (trEMBLrel. 24, Created)
 DT 01-JUN-2003 (trEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (trEMBLrel. 25, Last annotation update)
 DE Cyclin B3.
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Testis;
 RA Lozano J.C., Schatt P., Picard A.;
 RT "Cloning of two mRNA coding for two isoforms of mouse cyclin B3.";
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ555464; CAD88194.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:000074; P:regulation of cell cycle; IEA.
 DR InterPro; IPR006670; Cyclin.
 DR InterPro; IPR004367; Cyclin_Cterm.
 DR InterPro; IPR006671; Cyclin_N.
 DR Pfam; PF00134; cyclin; 1.
 DR Pfam; PF02384; cyclin C; 1.
 DR SMART; SM00385; CYCLIN; 2.
 KW Cyclin.
 SQ SEQUENCE 1396 AA; 158968 MW; 878PB226546F7FC0 CRC64;

 Alignment Scores:
 Pred. No.: 0.000117 Length: 1396
 Score: 170.50 Matches: 185
 Percent Similarity: 31.51% Conservative: 109
 Best Local Similarity: 19.83% Mismatches: 344
 Query Match: 2.22% Indels: 295
 DB: 11 Gaps: 37

 US-09-874-162A-7 (1-4409) x Q810T3 (1-1396)

 QY 106 TGGCGACTCCACTTCCACC-----CTGGCGCACTCATCGAGCAC 147
 DB 160 CysAlaPheHisGlnGluThrLeuLeuMetGluLysProLeuLeuValGluThr 179

 QY 148 ATCGAGGACCAACACATGATACAGATCCACGGTTTGTAGAAAAACAAGATTA---CAG 204
 DB 180 GluAspTyrAsnGluPheAspThrGluLeuMetThrSerLysLysAspLysProGlu 199

 QY 205 CAGCAACCTATGTTGCCCTGAGTTACATAAATAGATTCATGACAGATGCTGCCGCCGA 264
 DB 200 AspProThrIleIle-----Glu 205

 QY 265 GAGCAGGAGTCCCTAAAGAAGAGATTGAGCGAAGCTCTGCTGACTCTGTCC----- 318
 DB 206 GluMetThrAspLeuLysLysSerValIleArgLysValThrLeuThrSerProLeu 225

 QY 319 -----AGCTCA 324
 DB 226 TrpLeuLysAsnLysHisValValGlnGluLysProValIleGlnLysSerSer 245

 QY 325 GTGTCTCGA-----GGGAATGTCTCCACTCCGCCAGCCACAGCAGTGGAGC 372
 DB 246 PheLysArgIleSerLeuValSerAsnValValThrThrLysGlu----- 260

 QY 373 CTATCTCCCGCGTACGCCACCCATACCCCTCTCTTATCTCCGAGCAGCAGCTCCG 432
 DB 261 ---LysProProValLysLysPro-----HisPheArgLysLysLysPro 274

 QY 433 ACAGAGCCCAACACAGATCTATAGATTCTTGAATCGGAATCTCATAGCACCATATTT 492
 DB 275 Thr-----ThrGluMetLysSerLeuLeuGlnGluProSerLeu----- 287

 QY 493 TTGCACAGAACTCTTACTTACATGTCTCATCGAAACTCCAGAAACATCAAAAGGAAA 552
 DB 288 -----GluGluLysTyrAsn 292

 QY 553 ACATTAAAGTTGATGATGATTATCAMAAGTAGAGAAAANTGAAGGAGAGCAGAATCT 612
 DB 293 ThrGlnGluAspAlaSerIleLeuLysLysProGlnValLeuGlnGlnAsnThrAsn 312

 QY 613 CATAGCTCTCAGCTCAATTGCGAGTTACGTTTACTGTTTCTTCCACAAAATGAT--- 669
 DB 313 LysAspAlaThrLeuThrGluProValThrPheLysGlyLysHisSerAlaAsnGluAla 332

670 -----AAGCCATCAACCAAACTCAGAAAT-----GAACAAAATTCGTGT 708
 DB 333 ThrHisThrLysLysProSerSerSerLysAsnAsnProAspProGlnGlnLysGlyThr 352

 QY 709 ACCGTGAAGTCTCGTTGTG----- 729
 DB 353 AsnLeuArgProLeuArgValHisProValThrTyrGluAsnGluProMetSerSerLys 372

 QY 730 AAGATTTGCCACAAAAAAGA-----AAGGATGTA 759
 DB 373 LysSerThrThrLysLysLysAspSerHisPheHisGlyProSerValLeuProAspLys 392

 QY 760 AGTTGTCCAATAAGCAAGTCCACAGGTAAAGCAGGTGCTTTGATTCCTGACCTC 819
 DB 393 HisSerProGlnMetGluValSerThrValLysSerLeuAlaLeuProAsnProThr 412

 QY 820 AATCAACAACAAACCCGAAATTTCCGTCCTTCAGTTTCCAGTATTCAGTATGTAACCT 879
 DB 413 ThrGluGluLysMetLeuHisPheProValAlaThrValLeuGluLysGlnHisAsnMet 432

 QY 880 AGTAACAGCCATATGGTGAAGTCTTACTCGTTGCTATTAGAGTACTCGT---CCAGGA 936
 DB 433 GlyGluAlaProCysLysLysLysProSerProLeuArgLysGlnGlnLeuProLys 452

 QY 937 AGAAGAGAGTTTAAATGAATGATTATGAGAAACCAACCAATGAAATAT--- 984
 DB 453 ArgArgArgPhe-----PheSerAsnSerAlaValGlnGluThrValIleArgLysPro 470

 QY 985 -----GATGTCAATGAAGAGCTTCCA---GCCAGA 1011
 DB 471 LeuPhePheLysMetSerThrThrGluLysAspProSerGlnTrpProSerAlaLeu 490

 QY 1012 AGAAACGAAATCGTGAAGTGGGAAAGACATTTGTCACAAATGACAGATTTGAT 1071
 DB 491 ProLysLysHisIleSerProGlyGluLeuSerLysGlnLysLysGlnHisValSerPro 510

 QY 1072 AAAAAACGCGCTACAGCTTTAGATGGGGAATATCAAGTCCCATGCAAGGAAATCGAA 1131
 DB 511 LysHisAsn-----MetGluGluAspSer 518

 QY 1132 GAATGTCCAATAAGCAAGAAAGAGCAACATGGGAGACTATTCTTGATGGGAAGAGCTG 1191
 DB 519 GlnCys-----TrpLeuAspSerAlaPheLysLysGlnLeu 530

 QY 1192 CTCCCATCAAAACATTTCTCAGGACCTACCTGCGAGTTCCTCTCTCTGTTGG----- 1245
 DB 531 SerArgGluGluProAlaSerThrHisThrProLeuLysLeuGluMetGlnGlnAlaIle 550

 QY 1246 ACAGGAGAGACCAATGATAAATCTACGGCTCTATTGCGCAACCTCTTGCCACTAGAAAT 1305
 DB 551 ThrLysGluThrGlyPheHisLeuArgAsnProLeuValLeuProThrValThrSerGlu 570

 QY 1306 TCAGAGAGTCTCCATCAGGAAAAACAAGCTGGTTTCACTTAAACCTACTCAAACTATTGCT 1365
 DB 571 AlaLysSerLeuThrLysGlu-----ProProSerPheArgGluGlnAsnThrSerLeu 588

 QY 1366 GTTAAAGAATCATTTGACTACAGATCTACAAACAGAAAGAAAGGATACT----- 1416
 DB 589 LeuLysArgLysSerThrThrHisThrIleThrLeuGlnGlnAlaGlnSerGluTrpGln 608

 QY 1416 ----- 1416
 DB 609 GluMetThrAspGluAspArgAsnLeuPheSerIleLysProGlySerHisArgLysGlu 628

 QY 1417 -----CCAATGAAAAAC-----CGACAA 1434
 DB 629 ProIleProGluPheLeuGlnAsnProLeuProAsnGluAsnCysLeuIleSerGln 648

 QY 1435 AAATTAAAGATATTTTATCAGTTTCTCTATAACAACAATACAGGCAACAACTGAAGCA 1494
 DB 649 LysLeuSerHisSerMetProPheAlaSerGlnLysThrThrSerGlnGluArgAlaHis 668

 QY 1495 AGA-----GATGACCTGCATTGC 1512

Job time : 625 secs

```

|||
Db 669 ArgLysGluSerValAlaSerAsnAspLysAsnPhePheSerGlnAspLeuPheSer 688
|||
QY 1513 CTTGGTGTACTCTGAAGTGGCAAACTTTATAGTTTACTCAAGCATCTTAAACTCTGC 1572
|||
Db 689 ProPheSerSerAlaAspGluAspThrLeu----- 698
|||
QY 1573 CATACGAGATTATCTTCAACTATGTTTATCATCAAAAGGTGCTAGATAGATGTTTCT 1632
|||
Db 699 -----LysPheHisLysSerLeuAspPheGlnGluValAspArgLysAsnAspSer 716
|||
QY 1633 ATCAATGAGTGTATGATGCTCTATCGAGAAATCTCAGGATATTCATCGCCAACT 1692
|||
Db 717 HisLysLysMetPheAspSerGlnAspSerValSerGluGluGlu----- 731
|||
QY 1693 GGATTGCTTTTAGTCGCAACGGACAGTTAAGAGAACACCTATCACATATTCCTTGTG 1752
|||
Db 732 -----SerPheLeuArgLys-----LeuPhe 738
|||
QY 1753 TCCAGGCCAAAA-----CGAACAAAAGCAAGCATG 1782
|||
Db 739 CysLysAspArgCysSerSerThrGluLeuSerGlnGluArgThrValAlaLeuGlu 758
|||
QY 1783 TCTGAATTTCTT-----GAATCTGAAGATGGGAAGTAGAACACAA 1824
|||
Db 759 GlnGluPheLeuLeuLeuLysIleLeuAsnGluAsnThrSerSerAspValAspGluPro 778
|||
QY 1825 AGACATATAGTAGTGGCCAACTGCTGTATTTCCATAGTATGATGATGCTCTACCTCTC 1884
|||
Db 779 LeuSerHisGlnSerProHis-----IleGlnAsnHisSerAspThrThrLysGluAla 796
|||
QY 1885 -----CGTCCACAAGAAATGAGAA-----GTA 1905
|||
Db 797 LeuGluAlaSerGluAlaLeuGluAlaProGluAlaLeuGluThrLeuGluAlaLeuVal 816
|||
QY 1906 GATAGTGAAGATGAAAGATCCTGATGCGTAAAGTAAAGAAAA----- 1947
|||
Db 817 AlaSerGluAspLeuGluGluProLeuAsnIleLeuGluLeuLeuSerThrGluAsnMet 836
|||
QY 1948 -----ACCATTAACAATGAAGAGTTTCTGATGTTTAAATGAGGA 1989
|||
Db 837 ValAlaLeuMetLysMetLeuValThrGluAspGluSerThrLysAspSerPheSerGly 856
|||
QY 1990 GAGAAAGAGTGAAGAACTCTGGAATCTCCATGTCATGAAGCATGGGTTTATTCGTGAC 2049
|||
Db 857 AsnTyrThrAlaAlaArg-----GluAlaHisAlaGluLysSerLeuSerLeuGluGlu 874
|||
QY 2050 AATCAATGAATCATGCTGCTGATGCTGTTGTAGAAAATATGACACAGAAAATAATTAA 2109
|||
Db 875 ThrSerIleAsnGluAlaAlaThrLeuLysGluSerLeuSerSerGlnGluLysHisArg 894
|||
QY 2110 AAGAAATTTATGCGAAACTTCATGCTTCATCTAGTCAGCATGACATTTAATCTTTATT 2169
|||
Db 895 AlaGluLeuValThrValLeuLysGluLeuValLeuMetLysAsnProSerLeuLys 914
|||
QY 2170 AGCATAATG-----TCAATAGATAAGCTGTACCAAGCTCCGTGAA 2211
|||
Db 915 ArgValAlaLeuAlaPheGlnGluAsnProSerAsnAsnValGluThrLeuLeuArgGlu 934
|||
QY 2212 ATGCAGCAAAAATAGAAAGGGGAATCTGCTTCCCTGCAAAACGAAAGAAATAACTGAA 2271
|||
Db 935 ValLeuAlaLeuValGluAsn-----SerThrAlaAspGluSerThrLeuGln 950
|||
QY 2272 GAACAAAATGGACACAGCAAAAT-----GCATTAGTGAAT 2307
|||
Db 951 GluLysProSerThrLysThrAspValThrProLysGluLeuLeuAlaLeuGluGluAsn 970
|||
QY 2308 AACTCAAAAGAGAAAGCTTTGGAAACAGATAGTGTCTCA 2346
|||
Db 971 SerSerAsnLysLysAlaAsnProMetAspSerLeuSer 983
|||
```